Chapter 1

Introduction

Overview of concepts in biomolecular modelling

The basic challenge of biomolecular modelling is to describe the properties and behaviour of such systems in terms of interactions between specified sites. The interaction energy or potential energy of the system is generally assumed to be a more or less complicated function

$$V(x_1, x_2, x_3 ... x_N) \quad = \quad V(\vec{x}) \qquad (1.1)$$

of (Cartesian) coordinates x_i of all interaction centres.

The interaction function (1.1) is an effective interaction: it describe the interaction between explicitly treated degrees of freedoms averaged over the omitted degrees of freedoms.

Generally, modelling of a biomolecular system involves three choices:

1. Which degrees of freedom are treated explicitly in (1.1)

2. Which interaction function or force field $V(\vec{x})$ is used to calculate the potential energy of the system and the forces along the explicitly treated degrees of freedom.

3. Which algorithm is used to sample or simulate the motion along the degrees of freedom, or to search for those parts of configurational space for which $V(\vec{x})$ is minimal, that is, the regions of low energy and interest.

Choices one and two are highly dependent on each other and will therefore be treated together in the next section.

1.1 Force fields in biomolecular modelling

Based on the kind of question to answer, different kind of interaction functions are used. Historically, the first interaction functions in biomolecular modelling were empirical potential energy functions on an atomistic level and meant to be based on and reproduce real physical energies [1]. Expressed in terms of classical empirical functions, the potential energy is the sum of several components in which the mathematical functional forms of each contribution is based largely on phenomenological concepts concerning the nature of each energy term and easy to compute formulations. As an example for many force fields of this type [2–12] the functional forms of the GROMOS87 force field [12] are shown.

$$V = \sum_{n=1}^{N_b} \frac{1}{2} k_n^b \left(d_n - d_n^0\right)^2 \tag{1.2}$$

where the right side contributes for N_b bond interactions when d is the actual bond length, d_0 the equilibrium distance of the bond and k^b the force constant for the harmonic interaction. Similar forms are used for bond angle energy contribution

$$+ \sum_{n=1}^{N_\theta} \frac{1}{2} k_n^\theta \left(\theta_n - \theta_n^0\right)^2 \tag{1.3}$$

where θ is the bond angle and θ^0 is the ideal value. k^θ is the force constant of the bond angle stretching. A so–called improper dihedral angle energy contribution is used, to retain planarity and chiral centres. The improper torsion angle ξ is harmonically restrained to the value ξ^0 with a force constant k^ξ.

$$+ \sum_{n=1}^{N_\xi} \frac{1}{2} k_n^\xi \left(\xi_n - \xi_n^0\right)^2 \tag{1.4}$$

Further energy terms are dihedral angle potential energies with periodicity η and energy maximum at $\phi = \delta$

$$+\sum_{n=1}^{N_\phi} k_n^\phi \left(1 + \cos(\eta_n\ \phi_n - \delta_n)\right) \tag{1.5}$$

and Lennard–Jones and Coulomb energy contributions of all non–bonded atom pairs with the respective interaction parameters A and B, atomic charges q and vacuum and relative dielectric constants ϵ_0, respectively ϵ_d.

$$+\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{A_{i,j}}{d_{i,j}^{12}} - \frac{B_{i,j}}{d_{i,j}^{6}} - \frac{q_i q_j}{4\pi\epsilon_0\epsilon_d\ d_{i,j}} \tag{1.6}$$

The parameters describing the interaction properties (d^0, k^b, θ^0, k^θ, ξ^0, k^ξ, η, δ, k^ϕ, A, B, q) may be obtained by fitting to experimental structural, spectroscopic and thermodynamic measurements on small organic molecules. In addition, information from quantum mechanics calculations is commonly used in the parametrisation process when no or very poor experimental data are available, like for atomic charge distribution or transition state energies.

For the sake of speed, less interesting degrees of freedom are omitted and included implicitly in the calculations. Most obvious are electronic degrees of freedom which are not treated explicitly. Solvent molecules surrounding the bio–molecule may be ignored or considered only in a crude mean manner. Interactions of non–polar hydrogen atoms are included, in an average sense, into heavy atom interactions which not only reduces the number of interaction sites significantly, but also allows bigger time steps to resolve movements of the system.

To increase the accuracy of empirical molecular mechanics force fields, some of these and additional degrees of freedoms are considered in more recent force fields. The use of all atoms of the system and a box of solvent molecules as environment has become fairly standard in molecular simulations and even electronic degrees of freedoms have been included for parts of systems to simulate biomolecular reactions [14].

In addition to treating more degrees of freedom explicitly, more sophisticated functional forms are used account for coupling of different energy contributions. For example the DISCOVER force field contains a term coupling bond length and angles.

$$V^{b,\theta} = \sum_{i=1}^{N_b} \sum_{j=1}^{N_\theta} k^{b,\theta} \left(d_i - d_i^0\right) \left(\theta_j - \theta_j^0\right) \tag{1.7}$$

These *state of the art* force fields are able to reproduce a variety of physico–chemical properties with high accuracy for a big variety of molecules [15].

A drawback of these physically based force fields is that energy and forces are rather expensive to calculate and large problems in structural molecular biology are very expensive or impossible to solve.

Thus, another type of interaction functions has become fashionable. These are problem specific interaction function which typically drastically reduce the number of explicitly treated degrees of freedom by using simplified representations of molecules. Typically, these simplified representations group atoms into a single interaction centre and the interactions are described by mean interactions. As a matter of fact these interaction functions are not necessarily physically based and it is only appropriate to use them in the context they are designed for. Nevertheless, due to the simplicity and superiority over general physical force fields within this special context they are useful tools to attack special problems like long time simulations [16], protein folding [17], inverse protein folding [18] and fold recognition [19], to mention a few.

1.2 Methods to search and sample conformational space

The previous section describes energy functions used in molecular modelling and this section gives an overview of sampling and search methods. We strictly separate this section from the previous one since sampling and searching methods can generally be applied to any energy or quality function. For convenience, one can further classify the methods into sampling techniques (which aim to generate a known distribution) and pure search methods which may not generate a defined distribution.

1.2.1 Sampling methods

Under sampling methods one summarises techniques to sample a quality function under the constraint that the resulting samples obey a certain distribution and assuming a sufficient number of samples.

Brownian dynamics sampling

The most trivial case is an uniform distribution function of samples over variables. The kind of simulations to achieve such a distribution are known as *random walk* or Brownian dynamics simulations in which the change of variables are random throughout the simulation [20]. A more interesting case, however, is the generation of Boltzmann–distributed samples, since this distribution is the statistical–mechanical distribution of energies in molecular systems. In other words, molecules naturally adopt conformations reflecting this kind of energy distribution. The probability $p(\vec{x})$ of each sample is given by

$$p(\vec{x}) = \exp\Big(-\frac{V(\vec{x})}{k_B T}\Big) \tag{1.8}$$

where k_B is Boltzmann's constant and T is the (absolute) temperature. From this probability function it is obvious that regions of low energy on the potential energy surface are more important than high energy regions and a biased sampling has to be performed to obtain the desired ensemble.

Metropolis Monte Carlo sampling

The Metropolis Monte Carlo algorithm [21, 22] is a method which applies a random sequence of perturbations to the configuration of variables from the previous step. The new configuration is accepted with a probability dependent on the change in the energy function.

$$p = \exp\Big(-\frac{V(\vec{x}) - V_{old}(\vec{x})}{k_B T}\Big) \tag{1.9}$$

Although the resulting distribution is not dependent on the perturbation applied to the variables a proper choice of the size and type of perturbation is essential for efficient sampling. The Monte Carlo method usually performs well on smooth energy surfaces, but tends to be inefficient in cases where the energy surface is rough and highly anisotropic. Unfortunately, energy surfaces of bio–molecules are usually of the second type.

Molecular dynamics sampling

Molecular dynamics simulations are based on Newton's equation of motion describing dynamics of classical particles. Given the potential energy function $V(\vec{x})$, masses of the particles m and initial conditions of velocities $\vec{v}$ and coordinates $\vec{x}$ the evolution of the system in time (trajectory) is fully described by three differential equations.

$$\vec{F} = m\vec{a} \tag{1.10}$$

$$\dot{\vec{v}} = \vec{a} \tag{1.11}$$

$$\dot{\vec{x}} = \vec{v} \tag{1.12}$$

where forces are given are defined as the opposite of the first derivative of the potential energy function with respect to coordinates.

$$\vec{F} = -\frac{\partial V(\vec{x})}{\partial \vec{x}} \tag{1.13}$$

Since, for many particle systems no analytical solution of this set of equations exists, approximations are obtained by numerical integration. The most widely used numerical integrator in the field of molecular dynamics simulations is the Verlet algorithm and a modification of it: the leap–frog algorithm. Although they are relatively simple integrators, their big advantage is the necessity of only a single function evaluation per integration step. Using the leap–frog algorithm for integration equations (1.11) and (1.12) over a time step dt the equation of motions are

$$\vec{v}(t + dt/2) = \vec{v}(t - dt/2) + \frac{\vec{F}(t)}{m}\, dt \tag{1.14}$$

$$\vec{x}(t + dt) = \vec{x}(t) + \vec{v}(t + dt/2)\, dt \tag{1.15}$$

A feature of Newtonian dynamics is conservation of total energy which is the sum of potential $V(\vec{x})$ and kinetic energy $K(\vec{v})$

$$K(\vec{v}) = \sum_{n=1}^{N} \frac{1}{2}\, m_n\, (\vec{v}_n)^2 \tag{1.16}$$

Total energy is distributed to each degree of freedom statistically. As a consequence the probability of an energy state decreases exponentially with the amount of energy. When Newton's equations of motion are integrated, the system samples points on the potential energy surface within the constraint of constant total energy. Areas with higher potential energy than (constant) total energy are omitted from the sampling and in a strict sense molecular dynamics does not result in a Boltzmann ensemble. Nevertheless the error made is negligible for systems with a couple of degrees of freedoms in which kinetic energy can be freely distributed.

Often it is a difficult or even impossible task to find initial conditions with the desired total energy so thermostats are used to control temperature. These do not necessarily conserve total energy. A typical method is Berendsen's weak coupling method [23] which uses the coupling equation

$$\frac{dT(t)}{dt} = \tau_T^{-1} \left(T_0 - T(t) \right) \tag{1.17}$$

to force the actual simulation temperature $T(t)$ at time t toward the desired value T_0 with coupling constant τ_T. A small coupling constant (strong coupling) may distort the distribution of energies, but weak coupling does relatively little damage to the Boltzmann distribution.

In the same vein, other thermodynamic or experimental properties can be included in the equations of motion to force a system towards the reference state. Coupling to pressure [23], chemical potential [24], NMR distance measures [25] have been described.

1.2.2 Search methods:

In contrast to sampling methods, search methods are not supposed to generate ensembles with a known distribution of configurations. The only aim of a search method is to search efficiently a quality function for high quality regions. A variety of search methods are available, each with its particular strengths and weaknesses, depending on the form of the function $V(\vec{x})$ to search and the number and type of degrees of freedom of the system. Two basic types of search methods can be distinguished.

- *Systematic or exhaustive search methods*, which scan complete or significant part of the configurational space of the system. Certain subspaces can be excluded without loss of quality of the solution found, thanks to rigorous arguments that these subspaces cannot contain the desired solution. Such arguments are based on *a priori* knowledge, often of physical or chemical nature, about the structure of the space to be searched. Systematic search methods can only be applied to systems with few degrees of freedom, like small molecules [26], because of the exponential growth of the required computing effort as a function of the number of degrees of freedom.

- *Heuristic search methods*, which, although visiting a tiny fraction of the solution space, aim at generating a set of important configurations. These methods may general be divided into three types:

  1. *Non–step methods*, such as the so–called distance geometry metric matrix method [27, 28] which generate, for a limited set of problems which

can be cast into a distance based form, at least in principle, an uncorrelated set of random configurations.

2. *Step methods*, such as Monte Carlo [21,22] molecular dynamics, stochastic dynamics [29] or minimising based methods, that get a new configuration of the complete system from the previous one. A further distinction can be made according to the way the step direction and size is chosen:
   (a) according to the function value $V(\vec{x})$,
   (b) according to the gradient of the function $\nabla V(\vec{x})$,
   (c) at random,
   (d) according to a memory of the path followed so far.
3. Step methods, that build a complete configuration from fragments in a step–wise manner. Examples in molecular conformation search are build–up procedure [30,31], combinatorical build–up methods which make use of dynamic programming methods [32] and Monte Carlo chain growing methods [33].

The efficacy of search methods is tied to the nature of the function surface $V(\vec{x})$ that is to be explored to find high quality regions. Due to the occurrence of a multitude of low quality barriers separating high quality regions, the radius of convergence of step methods is generally very small. Therefore, a number of of techniques have been developed to enhance the searching power of searching methods. Two general types of *search enhancement techniques* can be distinguished.

1. *Deformation of the quality function surface in order to reduce barriers.*
2. *Scaling system parameters in order to improve the statistics of sampling.*

These are briefly reviewed in the next sections.

Deformation of the target function in order to reduce barriers

Enhanced efficient searching for high quality regions of $V(\vec{x})$ can be obtained by deforming the target function $V(\vec{x})$ in such a manner that barrier to conformational transitions are lowered or circumvented. A variety of procedures has been proposed, amongst which three basic types can be distinguished.

1. Generally, a *smoothing of the target function* $V(\vec{x})$ allows for a faster search for its minima. This technique has been applied to different problems such as molecular conformational search [34], protein folding [35], free energy computation [36] and structure determination based on nuclear magnetic resonance (NMR) data [37]. In NMR structure refinement the target function $V(\vec{x})$ includes a term representing atom pair distance restraints that can be smoothed by using only distance restraints which are local with respect to the covalent structure. This is the basis of the variable target function method [38] in which gradually longer–range atom pair interactions are included while minimising the energy function of the system. In X–ray crystallographic structure refinement the energy function can be made smoother by initially using only lower resolution reflections leading to an enlarged radius of convergence [39, 40]. Simplified non–bonded interaction functions used in NMR refinement also have fewer minima which leads to a larger radius of convergence [37].

2. *Incorporation of information on the energy hyper surface obtained during the search of the target function domain* is another possibility to enhance sampling. Once a local energy minimum is found, it is removed from the energy surface by a suitable local deformation of the potential energy function. This idea is the basis of the deflation method [41] and the local elevation search method [42]. Another way to introduce a memory into the search is the use of a potential energy term which is a running average over the trajectory or ensemble generated so far [43].

3. Barriers in the energy surface can be circumvented by an *extension of the dimensionality of the configurational space*. The technique of energy embedding locates a low energy configuration in high–dimensional space and gradually projects this to three–dimensional space while perturbing its energy and configuration as little as possible [46]. Variations on the original procedure have been proposed [47–49]. Dynamic search methods can also be used in conjunction with an extension of the dimensionality. By performing molecular dynamics in four-dimensional Cartesian coordinates, energy barriers in three-dimensional space can be circumvented [50] and free energy changes calculated [51].

Scaling of system parameters in order to improve sampling

The techniques discussed in the previous section all involved a deformation of $V(\vec{x})$ with the aim of facilitating the search for low energy configurations. Here, we list a few techniques that involve scaling of system parameters to enhance sampling.

1. The technique of *simulated temperature annealing* [52] involves simulation or search at high temperature T followed by gradual lowering of T. By raising the temperature the system may more easily surmount energy barriers, so a larger part of configurational space may be searched. The technique of simulated temperature annealing has been widely used in combination with Monte Carlo, molecular dynamics and stochastic dynamics [37, 53]. The potential energy annealing conformational search (PEACS) method [54] is based on a similar idea, but instead of temperature, potential energy is raised and annealed.

2. *Scaling of atomic masses.* In the classical partition function, the integration over the momenta of atoms can be carried out analytically, separately from that over the coordinates, if no constraints are applied. In that case, the atomic masses do not appear in the configurational integral, which means that the equilibrium properties will be independent of atomic masses. This freedom can be exploited in different ways to enhance the sampling. By increasing the mass of specific parts of a molecule its relative inertia is enhanced, which eases the surmounting of energy barriers [55], and may allow for larger time steps.

3. *Enhanced sampling by a mean field approximation* is obtained by separating molecular systems into two parts, A and B, each of which moves in the average field of the other. The initial configuration of the system consists of N_A identical copies of part A and N_B identical copies of part B, where the positions of corresponding atoms in identical copies may be chosen to be identical. The force on atoms in each copy of part A exerted by the atoms in all copies of part B is scaled by a factor N_B^{-1}, in order to obtain the the mean force exerted by part B on the individual atoms of part A. The force on atoms in each copy of part B exerted by the atoms in all copies of part A is scaled by a factor N_A^{-1}, in order to obtain the the mean force exerted by part A on the individual atoms of part B. The forces between the different copies of A are zero, and so are the forces between different copies of part B. The MD simulation

involves the integration of Newton's equation of motion, $\vec{F} = m\vec{a}$ for all copies of part A and B simultaneously. Thus, one obtains N_A individual trajectories of part A in the mean field of part B and vice versa. This comes at the loss of correct dynamics: Newton's third law, $\vec{F}_{AB} = -\vec{F}_{BA}$, is violated. The technique only enhances efficiency when the system is partitioned into parts of very different size, e.g. $size(\mathrm{A}) << size(\mathrm{B})$ and the bigger part is represented by one copy: $N_B = 1$. Enhanced searching and sampling procedures based on a mean field approximation have been proposed in different forms [56–65] and have been applied to the diffusion of CO molecules in the field of a protein molecule [58], to the conformational equilibrium of protein side chains [59,61], to the determination of protein loop conformations [62], and to searching for the minimum energy conformation of polypeptides [63–65].

1.3 Summary of the following chapters

The following chapters describe and assess novel function optimisation methods in different contexts of molecular modelling. Chapters 2, 3 and 4 describe methods to optimise molecular geometries by perturbing the potential energy surface. Chapter 5 describes a special purpose low resolution force field and shows how optimisation methods can be applied to the force field itself.

In Chapter 2, a method searching conformational space based on the history of the trajectory is presented. In heuristic search methods, such as MD, a particular region of conformational space is often repetitively visited, thereby reducing the rate of discovery of other low energy regions. The method presented overcomes this limitation by elevating the energy surface in regions already visited by the molecule. Thus, a more efficient search of conformational space can be deduced.

In Chapter 3, a method to search and sample conformational space by taking advantage of a mean field description of (parts of) a molecule is described. A mean field is the average field of a Boltzmann weighted ensemble of conformations. This means that the weight (probability) of each conformation has to be introduced into the simulation. Although the probabilities as extra variables have to be optimised in addition to the coordinates, the searching power is improved. The strength of the mean field approach in conformational sampling is demonstrated and the importance of additional weighting of the ensemble in the calculation is outlined.

In Chapter 4, a method to search for the global energy minimum conformation using analytical operator equations to deform the potential energy surface is presented. Searching for the global energy conformation of molecules is complicated by the rough landscape of the energy hyper surface. When this roughness is smoothed, a molecule is more likely find its global energy minima in the funnel like energy surface. The method and its performance compared to a standard optimisation method is presented.

In Chapter 5, a low resolution force field specialised for the fold recognition problem is designed and an optimal set of parameters found. The problem of fold recognition is to recognise whether a sequence is likely to fold into a particular structure. To express the likelihood a score or energy is calculated based on the sequence, coordinates and function parameters. In this chapter the force field parameterisation process is cast into an optimisation problem. This problem is solved and the results are evaluated.

References

[1] McCammon, J.A., Gelin, B.R. and Karplus, M., Dynamics of folded proteins, *Nature* **267**, 585–590 (1977).

[2] Levitt, M. and Lifson, S., Refinement of protein conformations using a macromolecular energy minimization procedure, *J. Mol. Biol.* **46**, 269–?? (1969).

[3] Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S.Jr. and Weiner, P., A new force field for molecular mechanics simulations of nucleic acids and proteins, *J. Am. Chem. Soc.* **106**, 765–784 (1984).

[4] Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., An all atom force field for simulations of proteins and nucleic acids, *J. Comput. Chem.* **7**, 230–252 (1986).

[5] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., CHARMM: A program for macromolecular energy, minimization and dynamics Calculations, *J. Comput. Chem.* **4**, 187–217 (1983).

[6] Biosym Technologies, 9685 Scraton, Road San Diego, CA 92121, *DISCOVER force field.*

[7] Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., Energy parameters in polypeptides. 7. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids, *J. Phys. Chem.* **79**, 2361–2381 (1975).

[8] NEMETHY, G., GIBSON, K.D., PALMER, K.A., YOON, C.N., PATERLINI, G., ZAGARI, A., RUMSEY, S. and SCHERAGA, H.A., Energy parameters in polypeptides. 10. Improved geometrical parameters, and nonbonded interactions for the use in the ECEPP/3 algorithm, with application to proline peptides, *J. Phys. Chem.* **96**, 6472–6487 (1992).

[9] LEVITT, M., Molecular dynamics of native protein. I. Computer simulation of trajectories, *J. Mol. Biol.* **168**, 595–617 (1983).

[10] LEVITT, M., Molecular dynamics of native protein. II. Analysis and nature of motion, *J. Mol. Biol.* **168**, 621–657 (1983).

[11] LEVITT, M., HIRSHBERG, M., SHARON, R. and DAGGETT, V., Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution, *Comp. Phys. Comm.* **91**, 215–231 (1995).

[12] VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C., *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, 1987.

[13] RYCKAERT, J.-P., CICCOTTI, G. and BERENDSEN, H.J.C., Numerical Integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* **23**, 327–341 (1977).

[14] FIELD, J.F., The simulation of chemical reactions, in *Computer simulation of biomolecular systems*, edited by VAN GUNSTEREN, W.F., WEINER, P.K. and WILKINSON, A.J., volume 2, pp. 82–123, ESCOM, Leiden, 1993.

[15] MAPLE, J.R., HWANG, M.-J., STOCKFISCH, T.P., DINUR, U., WALDMAN, M., EWING, C.S. and HAGLER, A.T., Derivation of class II force fields. 1. Methodology and quantum force field for the alkyl functional group and alkane molecules, *J. Comput. Chem* **15**, 162–182 (1994).

[16] GRUBMÜLLER, H. and TAVAN, P., Molecular dynamics of conformational substates for a simplified protein model, *J. Comput. Phys.* **??**, ?? (19).

[17] LEVITT, M. and WARSHEL, A., Computer simulation of protein folding, *Nature* **253**, 694–698 (1975).

[18] PABO, C., Designing proteins and peptides, *Nature* **301**, 200 (1983).

[19] HENDLICH, M., LACKNER, P., WEITCKUS, S., FLÖCKNER, H., FROSCHAUER, R., GOTTSBACHER, K., CASARI, G. and SIPPL, M.J., Identification of native protein folds amongs a large number of incorrect models: The calculation of low energy conformations from potentials of mean force, *J. Mol. Biol.* **216**, 167–180 (1990).

[20] ERMAK, D.L. and MCCAMMON, J.A, Brownian dynamics with hydrodynamic interactions, *J. Chem. Phys.* **69**, 1352–1360 (1977).

[21] METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. and TELLER, E., Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087–1092 (1953).

[22] FRENKEL, D., Monte Carlo Simulations: A primer, in *Computer simulation of biomolecular systems*, edited by VAN GUNSTEREN, W.F., WEINER, P.K. and WILKINSON, A.J., volume 2, pp. 27–66, ESCOM, Leiden, 1993.

[23] BERENDSEN, H.J.C., POSTMA, J.P.M., VAN GUNSTEREN, W.F., DINOLA, A. and HAAK, J.R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **81**, 3684–3690 (1984).

[24] BEUTLER, T.C. and VAN GUNSTEREN, W.F., Molecular dynamics simulations with first order coupling to a bath of constant chemical potential, *Mol. Sim.* **14**, 21–34 (1994).

[25] NANZER, A.P., HUBER, T., TORDA, A.E. and VAN GUNSTEREN, W.F., Molecular dynamics simulation using weak–coupling NOE distance restraining, *J. Biomol. NMR* **8**, 285–291 (1996).

[26] D.D.BEUSEN and MARSHALL, G.R., The generation of three-dimensional structures from NMR derived constraints, in *Protein Structure and Engineering*, edited by JARDETZKY, O., pp. 97–109, Plenum Press, N.Y., 1989.

[27] BLUMENTHAL, L.M., *Theory and applications of distance geometry*, Chelsea, New York, 1970.

[28] CRIPPEN, G.M., A novel approach to the calculation of conformation: Distance geometry, *J. Comput. Phys.* **26**, 449–?? (1977).

[29] VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C., A leap-frog algorithm for stochastic dynamics, *Mol. Simulation* **1**, 173–185 (1988).

[30] GIBSON, K.D. and SCHERAGA, H.A., Revised algorithm for the build–up procedure for predicting protein conformations by energy minimisation, *J. Comput. Chem.* **8**, 826–834 (1987).

[31] SCHERAGA, H.A., Searching conformational space, in *Computer simulation of biomolecular systems*, edited by VAN GUNSTEREN, W.F., WEINER, P.K. and WILKINSON, A.J., volume 2, pp. 231–248, ESCOM, Leiden, 1993.

[32] VAJDA, S. and DELISI, C., Determining minimum energy conformations of polypeptides by dynamic programming, *Biopolymers* **29**, 1755–1772 (1990).

[33] VELIKSON, B., GAREL, T., NIEL, J.-C., ORLAND, H. and SMITH, J.C., Conformational distribution of Heptaalanin: Analysis using a new Monte Carlo chain growth method, *J. Comput. Chem.* **13**, 1216–1233 (1992).

[34] PIELA, L., KOSTROWICKI, J. and SCHERAGA, H.A., The mulitiple minima problem in conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method, *J. Phys. Chem.* **93**, 3339–3346 (1989).

[35] AMARAN, P. and STRAUB, J. E., Folding model proteins using kinetic and thermodynamic annealing of classical density distribution, *J. Phys. Chem.* **99**, 14840–14853 (1995).

[36] MARK, A.E., VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C., Calculation of relative free energy via indirect pathways, *J. Chem. Phys.* **94**, 3808–3816 (1991).

[37] NILGES, M., CLORE, G.M. and GRONENBORN, A.M., Determination of three-dimensional structures of proteins from interproton distance data by hybrid dynamical simulated annealing from a random array of atoms, *FEBS Lett.* **239**, 129–136 (1988).

[38] BRAUN, W. and GO, N., Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *J. Mol. Biol.* **186**, 611–26 (1985).

[39] GROS, P., KALK, M. FUJINAGA B.W. DIJKSTRA K.H. and HOL, W.J.G., Crystallographic refinement by incorporation of molecular dynamics: Thermostable serin proteiase Thermitase complexed with Eglin C, *Acta Cryst.* **B45**, 488–499 (1989).

[40] GROS, P. and VAN GUNSTEREN, W.F., Crystallographic refinement and structure–factor time–averaging by molecular dynamics in the absence of a physical force field, *Mol. Sim.* **10**, 377–395 (1993).

[41] CRIPPEN, G.M. and SCHERAGA, H.A., Minimization of polypeptide energy, VIII. Application of the deflation technique to a dipeptide, *Proc. Natl. Acad. Sci. USA* **64**, 42–49 (1969).

[42] HUBER, T., TORDA, A.E. and VAN GUNSTEREN, W.F., Local elevation: A method for improving the searching properties of molecular dynamics simulation, *J. Computer Aided Mol. Design* **8**, 695–708 (1994).

[43] TORDA, A.E., SCHEEK, R.M. and VAN GUNSTEREN, W.F., Time-dependent distance restraints in molecular dynamics simulations, *Chem. Phys. Lett.* **157**, 289–294 (1989).

[44] SCHEEK, R.M., TORDA, A.E. and VAN GUNSTEREN, W.F., Time-averaged NOE distance restraints applied to Tendamistat, *J. Mol. Biol.* **214**, 223–235 (1990).

[45] GROS, P., VAN GUNSTEREN, W.F. and HOL, W.G.J., Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics, *Science* **249**, 1149–1152 (1990).

[46] CRIPPEN, G.M., Conformational analysis by energy embedding, *J. Comput. Chem.* **3**, 471–476 (1982).

[47] PURISIMA, E.O. and SCHERAGA, H.A., An approach to the multiple-minima problem by relaxing the dimensionality, *Proc. Natl. Acad. Sci. USA* **83**, 2782–2786 (1986).

[48] CRIPPEN, G.M., Linearized embedding: A new metric matrix algorithm for calculating molecular conformations subject to geometric restraints, *J. Comput. Chem.* **10**, 896–902 (1989).

[49] CRIPPEN, G.M. and HAVEL, T.F., Global energy minimization by rotational energy embedding, *J. Chem. Inf. Comput. Sci.* **30**, 222–227 (1990).

[50] VAN SCHAIK, R.C., BERENDSEN, H.J.C., TORDA, A.E. and VAN GUNSTEREN, W.F., A structure refinement method based on molecular dynamics in four spatial dimensions, *J. Mol. Biol.* **234**, 751–762 (1993).

[51] BEUTLER, T.C. and VAN GUNSTEREN, W.F., Molecular dynamics free energy calculation in four dimensions, *J. Chem. Phys.* **101**, 1417–1422 (1994).

[52] KIRKPATRICK, S., GELATT, C.D.JR. and VECCHI, M.P., Optimization by simulated annealing, *Science* **220**, 671–680 (1983).

[53] CAR, R. and PARRINELLO, M., A unified approach for molecular dynamics and density-functional theory, *Phys. Rev. Lett.* **55**, 2471–2474 (1985).

[54] VAN SCHAIK, R.C., VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C, Conformational search by potential annealing algorithm and application to Cyclosporin A, *J. Computer Aided Mol. Design* **6**, 97–112 (1992).

[55] MAO, B. and FRIEDMAN, A.R., Molecular dynamics simulation by atomic mass weighting, *Biophys. J* **58**, 803–805 (1990).

[56] GERBER, R.B., BUCH, V. and RATNER, M.A., Time dependent selfe consistent field approximation for intermolecular energy transfer. I. Formulation and application to dissociation of van der Waals molecules, *J. Chem. Phys.* **77**, 3022–3030 (1982).

[57] STRAUB, J.E. and KARPLUS, M., Energy equipartitioning in the classical time-dependent Hatree approximation, *J. Chem. Phys.* **94**, 6737–6739 (1990).

[58] ELBER, R. and KARPLUS, M., Enhanced sampling in molecular dynamics: Use of the time–dependent Hatree approximation for a simulation of carbon monoxide diffusion through Myoglobin, *J. Am. Chem. Soc.* **112**, 9161–9175 (1990).

[59] ROITBERG, A. and ELBER, R., Modelling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated an-

nealing methods to find minimum energy conformations, *J. Chem. Phys.* **95**, 9277–9287 (1991).

[60] AMARA, P., HSU, D. and STRAUB, J.E., Global energy minimum search using an approximate solution of the imaginary time Schrödinger equation, *J. Phys. Chem.* **97**, 6715–9721 (1993).

[61] HUBER, T., TORDA, A.E. and VAN GUNSTEREN, W.F., Optimization methods for conformational sampling using a Boltzmann-weighted mean field approach, *Biopolymers* **39**, 103–114 (1996).

[62] ZHENG, Q., ROSENFELD, R., VAJDA, S. and DILISI, C., Determining protein loop conformation using scaling-relaxation techniques, *Protein Science* **2**, 1242–1248 (1993).

[63] OLSZEWSKI, K.A., PIELA, L. and SCHERAGA, H.A., Mean field theory as a tool for intramolecular conformational optimization. 1. Test on terminally–blocked alanine and Met-Enkephalin, *J. Phys. Chem.* **96**, 4672–4676 (1992).

[64] OLSZEWSKI, K.A., PIELA, L. and SCHERAGA, H.A., Mean field theory as a tool for intramolecular conformational optimization. 2. Test on the homoploypeptides decaglycine and icosalanine, *J. Phys. Chem.* **97**, 260–266 (1993).

[65] OLSZEWSKI, K.A., PIELA, L. and SCHERAGA, H.A., Mean field theory as a tool for intramolecular conformational optimization. 3. Test on Melittin, *J. Phys. Chem.* **97**, 267–270 (1993).