Chapter 2

Protein-Protein Association

Biochemical processes depend on the ability of proteins to interact with each other, with nucleic acids, with lipids, with polysaccharides, and with substrate molecules. Specific molecular recognition is a prerequisite for antigen recognition by the immune system, for signal transduction in cells, for the adaption of cells to environmental conditions, and for many other physiological reactions (Stryer, 1990; Voet & Voet, 1995). Therefore, the understanding of protein association is required to gain insights into biochemical processes. Consequently, protein association is intensively explored experimentally as well as theoretically.

From the theoretical point of view, the docking problem consists of two tasks. First, it is required to find the correct binding site, and second, the correct binding affinity must be predicted. Theoretical studies investigate kinetic and thermodynamic properties of protein-protein and protein-ligand interactions. While free energy calculations and electrostatic calculations are able to investigate the thermodynamics of the binding reactions, the Brownian dynamics method is able to calculate the kinetics of association reactions if it is diffusion controlled. Many recent articles review this field of research (Janin & Chothia, 1990; Jiang & Kim, 1991; Madura *et al.*, 1994; Northrup, 1994; Kollman, 1994 Stanfield & Wilson, 1995; Ajay & Murcko, 1995; Gilson *et al.*, 1997a; Gilson *et al.*, 1997b; Stites, 1997; Sternberg *et al.*, 1998; McCammon, 1998).

A topic that is closely related to the docking of molecules to proteins is a field called *indirect drug design*. The goal of indirect drug design is to get a model of the binding pocket of a structurally unknown receptor. The crux is to align ligands that all bind to the same receptor (Hauswald, 1998). The aligned ligands allow conclusions about sites and structural motifs of the ligands that are important for the interaction with their receptor. The structural features of the ligands are assumed to present the ”negative” of the binding pocket at the receptor. While previous applications used the indirect-drug-design strategy only for small drug molecules, it is applied here to superimpose proteins that are able to perform the same physiological function but differ completely in structure (Ullmann *et al.*, 1997a; Ullmann *et al.*, 1998a). With the aid of the aligned protein structures, it is possible to propose which amino acids may have the same function in the investigated proteins.

In this chapter, I first introduce some basic principles on protein association, then I describe a method that I used during my PhD work to generate the docked complexes. A method to score the obtained docked complexes is also explained. Finally, I describe briefly the technique commonly used in indirect drug design that is applied here to superimpose and compare isofunctional proteins.

2.1 Thermodynamic Basis for Calculating Association Constants of Biomolecules

The association equilibrium of two molecules A and B to the complex AB is shown in eq 2.1.

$$\mathrm{A} + \mathrm{B} \overset{K_{\mathrm{AB}}}{\rightleftharpoons} \mathrm{AB}; \qquad K_{\mathrm{AB}} = \frac{[\mathrm{AB}]}{[\mathrm{A}][\mathrm{B}]} \tag{2.1}$$

where K_{AB} is the association constant and [A], [B], and [AB] represent the concentrations of A, B, and AB respectively. Association reactions of biomolecules are usually observed in an aqueous electrolyte. At equilibrium, the chemical potentials of the reactants and of the product state are equal (eq 2.2).

$$\mu_{sol,\mathrm{AB}} = \mu_{sol,\mathrm{A}} + \mu_{sol,\mathrm{B}} \tag{2.2}$$

The chemical potential of compound A is given by eq 2.3

$$\mu_{sol,\mathrm{A}} = \mu^{\mathrm{o}}_{sol,\mathrm{A}} + RT \ln \frac{\gamma_{\mathrm{A}}[\mathrm{A}]}{C^{\mathrm{o}}} \tag{2.3}$$

where $\mu^{\mathrm{o}}_{sol,\mathrm{A}}$ is the standard chemical potential, i. e., the chemical potential in the standard state, [A] is the concentration of compound A, C^{o} is the standard concentration in the same unit, γ_{A} is the activity coefficient of compound A, which approaches unity at low concentrations of the compound, R is the gas constant, and T is the temperature. The free energy of association $\Delta G^{\mathrm{o}}_{\mathrm{AB}}$ is obtained by combining eq 2.2 and eq 2.3.

$$\begin{aligned} \Delta G^{\mathrm{o}}_{\mathrm{AB}} &= \mu^{\mathrm{o}}_{sol,\mathrm{AB}} - \mu^{\mathrm{o}}_{sol,\mathrm{A}} - \mu^{\mathrm{o}}_{sol,\mathrm{B}} \\ &= -RT \ln \left(\frac{\gamma_{\mathrm{AB}}}{\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}} \frac{C^{\mathrm{o}}[AB]}{[A][B]} \right)_{eq} \\ &= -RT \ln K_{\mathrm{AB}} \end{aligned} \tag{2.4}$$

The standard chemical potential of a compound A in solution is given by eq 2.5.

$$\begin{aligned} \mu^{\mathrm{o}}_{sol,\mathrm{A}} &= \left(\frac{\partial G}{\partial n_{\mathrm{A}}} \right)_{T,p} \\ &= -RT \ln \left(\frac{1}{V_{N,\mathrm{A}} C^{\mathrm{o}}} \frac{Q_{N,\mathrm{A}}(V_{N,\mathrm{A}})}{Q_{N,0}(V_{N,0})} \right) + p^{\mathrm{o}} \overline{V}_{\mathrm{A}} \end{aligned} \tag{2.5}$$

Here, $Q_{N,\mathrm{A}}(V_{N,\mathrm{A}})$ is the canonical partition function for a system containing a large number of N solvent molecules and one solute molecule A at the volume $V_{N,\mathrm{A}}$. $V_{N,\mathrm{A}}$ is the volume, when the system is at standard pressure p^{o}. $Q_{N,0}(V_{N,0})$ is the canonical partition function for a system containing a large number of N solvent molecules without a solute. Now the volume of the system is $V_{N,0}$ at the same pressure p^{o}. The change of the equilibrium volume if one solute molecule is added to the system is $\overline{V}_{\mathrm{A}} = V_{N,\mathrm{A}} - V_{N,0}$. If $N \gg 1$, $\overline{V}_{\mathrm{A}}$ is the partial molar volume of the solute at infinite dilution in the solvent. The term $p^{\mathrm{o}}\overline{V}_{\mathrm{A}}$ describes the mechanical work associated with the volume change $\overline{V}_{\mathrm{A}}$, which is small at atmospheric pressure.

To find an expression for the ratio of the partition functions on the right-hand side of eq 2.5, an expression for the energy H of the system is required in terms of coordinates $(\mathbf{r} = (\mathbf{r}'_{\mathrm{A}}, \mathbf{r}_{\mathrm{S}}))$

and conjugated momenta ($\mathbf{p} = (\mathbf{p}_\mathrm{A}, \mathbf{p}_\mathrm{S})$) (eq 2.6).

$$\mathrm{H}(\mathbf{p}_\mathrm{A}, \mathbf{p}_\mathrm{S}, \mathbf{r}'_\mathrm{A}, \mathbf{r}_\mathrm{S}) = \sum_{i=1}^{M_\mathrm{A}, M_\mathrm{S}} \frac{\mathrm{p}_i{}^2}{2m_i} + U(\mathbf{r}'_\mathrm{A}, \mathbf{r}_\mathrm{S}) \tag{2.6}$$

Here, M_A and M_S are the numbers of atoms of the solute molecule and of the N solvent molecules respectively; $\mathbf{p}$ is the vector of the momenta of the $M_\mathrm{A} + M_\mathrm{S}$ atoms; $\mathrm{p}_i{}^2$ is the squared magnitude of the momentum of atom i; m_i is the mass of atom i; and $U(\mathbf{r}_\mathrm{A}, \mathbf{r}_\mathrm{S})$ is the potential energy, which depends on all coordinates of the system. For a system of classical particles, the expression in eq 2.7 is obtained,

$$\frac{Q_{N,\mathrm{A}}(V_{N,\mathrm{A}})}{Q_{N,0}(V_{N,0})} = \frac{\int\int e^{-H(\mathbf{p}_\mathrm{A}, \mathbf{p}_\mathrm{S}, \mathbf{r}'_\mathrm{A}, \mathbf{r}_\mathrm{S})/RT} \mathrm{d}\mathbf{p}_\mathrm{A} \mathrm{d}\mathbf{p}_\mathrm{S} \mathrm{d}\mathbf{r}'_\mathrm{A} \mathrm{d}\mathbf{r}_\mathrm{S}}{\sigma_\mathrm{A} \int\int e^{-H(\mathbf{p}_\mathrm{S}, \mathbf{r}_\mathrm{S})/RT} \mathrm{d}\mathbf{p}_\mathrm{S} \mathrm{d}\mathbf{r}_\mathrm{S}} \tag{2.7}$$

where σ_A is the symmetry number of the solute molecule. The symmetry number of the solvent molecule cancels. The integral over each momentum component extends from $-\infty$ to $+\infty$, and the position integrals range over all configurations that are consistent with molecules being intact and within their container.

To simplify eq 2.7, the molecular coordinate system is switched to allow the separation of the lab-frame coordinates of the solute atom $\mathbf{r}'_\mathrm{A}$ (external coordinates) from the internal coordinates of the solute. Three non-linear atoms are used to define this coordinate system. The first atom becomes the origin of the molecular coordinates. The vector from the first atom to the second atom defines the *x*-axis. The direction of the *y*-axis is given by the direction joining the first and the third atom, minus the *x*-component of that vector. The *z*-axis is given by the vector product of the *x*- and *y*-axis. The internal coordinates of the molecule then define the position of each atom. Since the first atom is in the origin of the molecular coordinate system, the second atom is on the *x*-axis of that coordinate system, and the third atom is in the *xy*-plane; these six coordinates correspond to the external molecular coordinates. The remaining set of $3M_\mathrm{A}-6$ coordinates are now denoted by $\mathbf{r}_\mathrm{A}$. The position of the origin of the molecular system, i. e., of the first atom, will be denoted by $\mathbf{R}_\mathrm{A}$. The three Eulerian angles, that define the orientation of the molecular frame relative to the lab-frame are $\xi_{\mathrm{A}1}$, $\xi_{\mathrm{A}2}$, and $\xi_{\mathrm{A}3}$. Thus the complete set of external coordinates is $\zeta_\mathrm{A} = (\mathbf{R}_\mathrm{A}, \xi_{\mathrm{A}1}, \xi_{\mathrm{A}2}, \xi_{\mathrm{A}3})$.

The integrals over the internal coordinates of the solute and over the solvent do not depend on the position and the orientation of the solute, i. e., on ζ_A. Thus this integral can be carried out leading to a factor of $8\pi^2 V_{N,\mathrm{A}}$. In the classical approximation, the integral over the momentum of each atom i contributes a factor of $(2\pi m_i RT)^{3/2}$. The integrals over the momentum of the solvent molecules cancel. Combining these expressions, the standard chemical potential is given by eq 2.8.

$$\mu^\mathrm{o}_{sol,\mathrm{A}} = -RT \ln \left(\frac{8\pi^2}{C^\mathrm{o} \sigma_\mathrm{A}} \prod_{i=1}^{M_\mathrm{A}} (2\pi m_i RT)^{3/2} \frac{Z_{N,\mathrm{A}}}{Z_{N,0}} \right) + p^\mathrm{o} \overline{V}_\mathrm{A} \tag{2.8}$$

$Z_{N,\mathrm{A}}$ and $Z_{N,0}$ are configuration integrals.

$$Z_{N,\mathrm{A}} = \int e^{-U(\mathbf{r}_\mathrm{A}, \mathbf{r}_\mathrm{S})/RT} \mathrm{d}\mathbf{r}_\mathrm{A} \mathrm{d}\mathbf{r}_\mathrm{S} \tag{2.9}$$

$$Z_{N,0} = \int e^{-U(\mathbf{r}_\mathrm{S})/RT} \mathrm{d}\mathbf{r}_\mathrm{S} \tag{2.10}$$

The definition of the standard chemical potential of the complex AB bears two problems. First, it is necessary to define the internal and external coordinates of the non-covalent complex AB. Second, only those configurations in which the two molecules are bound to each other should be considered. The first problem can be solved easily by defining the external coordinates of molecule B, ζ_B, relative to molecule A, so that the external coordinates of molecule B become internal coordinates of the complex. The second problem is of more general nature, especially for loosely-bound complexes. Mathematically, this problem might be formulated by including a step function $I(\zeta_B)$ into the configuration integral for the complex (eq 2.10), which equals unity for the complexed configurations and is zero otherwise.

$$\mu^{o}_{sol,AB} = -RT \ln \left(\frac{8\pi^2}{C^o \sigma_{AB}} \prod_{i=1}^{M_A+M_B} (2\pi m_i RT)^{3/2} \frac{Z_{N,AB}}{Z_{N,0}} \right) + p^o \overline{V}_{AB} \tag{2.11}$$

$$Z_{N,AB} = \int I(\zeta_B) J_{\zeta_B} \, e^{-U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)/RT} d\mathbf{r}_A d\mathbf{r}_B d\zeta_B d\mathbf{r}_S \tag{2.12}$$

Here, J_{ζ_B} is the Jacobian determinant for the Eulerian rotation of the molecule B relative to A.

However, the problem remains to decide at which boundaries the step function $I(\zeta_B)$ should be unity, i. e., how the complex is actually defined. If the binding between the two molecules A and B is tight, only the complexed configurations will contribute to the configuration integral significantly. In general, the potential of mean force should be negative in the range of ζ_B where the complex is formed. The potential of mean force, for instance obtained from a Brownian dynamics simulation (Thomasson *et al.*, 1997), can thus be used to define the bound state. The region in which the step function $I(\zeta_B)$ equals one should satisfy the following two conditions.

- The region should include all configurations that contribute significantly to the chemical potential of the complexed state.

- The region should not include a too large part of the phase volume of the uncomplexed configuration, since these do actually not contribute to $\mu^{o}_{sol,AB}$.

It is important to realize that the actual definition of the step function $I(\zeta_B)$ depends also on the experimental method, which has been used or should be used to determine the binding affinity.

Combining eqs 2.4, 2.8, and 2.11, we obtain an expression for the free energy of association ΔG^{o}_{AB}.

$$\Delta G^{o}_{AB} = -RT \ln \left(\frac{C^o}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \frac{Z_{N,AB} Z_{N,0}}{Z_{N,A} Z_{N,B}} \right) + p^o \Delta \overline{V}_{AB} \tag{2.13}$$

where $\Delta \overline{V}_{AB} = \overline{V}_{AB} - \overline{V}_A - \overline{V}_B$. The volume change $\Delta \overline{V}_{AB}$ is usually small at standard pressure, thus the associated mechanical work term is negligible. However, this contribution becomes important at high pressures ($\sim$100 at), which are found in deep ocean and in ultracentrifuge (van Eldik, 1993; Sindern *et al.*, 1995). Note that all mass-dependent terms cancel in eq 2.13.

2.2 Treatment of Solvent Molecules during Calculation of Association Constants

2.2.1 Explicit Solvent

An atomic representation of the solvent allows to model most solvent effects explicitly. This representation is often used in molecular dynamics and Monte Carlo simulations. Several water models such as TIP3P or SPCE exist (Jorgensen *et al.*, 1983). The force field parameters of these water models are chosen in order to reproduce physical properties of water. Changes in the free energies of association upon mutation and differences in association constants of different but similar ligands can be calculated by thermodynamic perturbation or integration methods (for review see van Gunsteren & Berendsen, 1990; Kollman, 1993; van Gunsteren *et al.*, 1993; Straatsma *et al.*, 1993; King, 1993). Since I did not use free energy calculations in my PhD work, I will not describe this method here.

2.2.2 Implicit Solvent

It is possible to separate the effect of solvent on the solute in a solvation energy term that depends on the conformation of the solute only. This separation is the theoretical basis for the use of implicit solvent models in calculations of thermodynamic properties of biological molecules. The standard free energy of binding in eq 2.13 contains three configuration integrals that involve the solute as well as the solvent: $Z_{N,\mathrm{AB}}$, $Z_{N,\mathrm{A}}$, and $Z_{N,\mathrm{B}}$. To achieve the separation of the solvent effect, the interaction of the solute with the solvent $\Delta U(\mathbf{r}_\mathrm{A},\mathbf{r}_\mathrm{S})$ is redefined as described in eq 2.14. The separation is demonstrated here for molecule A only, but can be also applied to B and AB.

$$\Delta U(\mathbf{r}_\mathrm{A},\mathbf{r}_\mathrm{S}) = U(\mathbf{r}_\mathrm{A},\mathbf{r}_\mathrm{S}) - U(\mathbf{r}_\mathrm{A}) - U(\mathbf{r}_\mathrm{S}) \tag{2.14}$$

Thus, the chemical potential of A becomes now

$$\mu^\mathrm{o}_{sol,\mathrm{A}} = -RT\ln\left(\frac{8\pi^2}{C^\mathrm{o}\sigma_\mathrm{A}}\prod_{i=1}^{M_\mathrm{A}}(2\pi m_i RT)^{3/2}Z_\mathrm{A}\right) + p^\mathrm{o}\overline{V}_\mathrm{A} \tag{2.15}$$

$$Z_\mathrm{A} = \frac{Z_{N,\mathrm{A}}}{Z_{N,0}} = \int e^{-(U(\mathbf{r}_\mathrm{A})+\mathcal{W}(\mathbf{r}_\mathrm{A}))/RT}\mathrm{d}\mathbf{r}_\mathrm{A} \tag{2.16}$$

where

$$\mathcal{W}(\mathbf{r}_\mathrm{A}) = -RT\ln\left(\frac{\int e^{-\Delta U(\mathbf{r}_\mathrm{A},\mathbf{r}_\mathrm{S})}e^{-U(\mathbf{r}_\mathrm{S})}\mathrm{d}\mathbf{r}_\mathrm{S}}{\int e^{-U(\mathbf{r}_\mathrm{S})}\mathrm{d}\mathbf{r}_\mathrm{S}}\right) \tag{2.17}$$

The argument of the logarithm in eq 2.17 is the pure-solvent average of the solute-solvent interaction potential for the conformation $\mathbf{r}_\mathrm{A}$ of molecule A. The standard free energy of association may now be rewritten by combining eq 2.13 and eq 2.16 as given by eq 2.18.

$$\Delta G^\mathrm{o}_\mathrm{AB} = -RT\ln\left(\frac{C^\mathrm{o}}{8\pi^2}\frac{\sigma_\mathrm{A}\sigma_\mathrm{B}}{\sigma_\mathrm{AB}}\frac{Z_\mathrm{AB}}{Z_\mathrm{A}Z_\mathrm{B}}\right) + p^\mathrm{o}\Delta\overline{V}_\mathrm{AB} \tag{2.18}$$

The Poisson-Boltzmann Equation (Hill, 1960), the Langevin-Dipole model (Warshel & Russel, 1984), and the solvent model proposed by Schaefer & Karplus, 1996 are examples for implicit solvent models.

2.2.3 Mixed Representation of the Solvent

A small number $n_A \ll N$ of solvent molecules that interact with the solutes A, B, and AB are represented explicitly, while the rest of the solvent molecules are treated by the solvation potential of mean force $\mathcal{W}$. If, together with the molecule A, a small fraction of n_A solvent molecules are represented explicitly, eq 2.5 changes to eq 2.19.

$$\begin{aligned}\mu^{o}_{sol,A} = & - RT\ln\left(\frac{1}{V_{N,A}C^{o}}\frac{Q_{N,A}}{Q_{N-n_A,0}}\right) + p^{o}(\overline{V}_{N,A} - \overline{V}_{N-n_A,0}) \\ & - RT\ln\left(\frac{Q_{N-n_A,0}}{Q_{N,0}}\right) + p^{o}(\overline{V}_{N-n_A,0} - \overline{V}_{N,0})\end{aligned} \tag{2.19}$$

The second line is directly related to the chemical potential of the solvent μ_S (eq 2.20), where μ_S is the actual chemical potential, *not* the standard chemical potential.

$$RT\ln\left(\frac{Q_{N-n_A,0}}{Q_{N,0}}\right) + p^{o}(\overline{V}_{N-n_A,0} - \overline{V}_{N,0}) = RT\sum_{i=1}^{n_A}\left(\ln\frac{Q_{N-i+1,0}}{Q_{N-i,0}}\right) + n_A p^{o}\overline{V}_S \tag{2.20}$$

Each term in the sum is the free energy upon removing one of the N solvent molecules from the pure solvent, $\overline{V}_S$ represents the partial molar volume of one solvent molecule. Eq 2.20 is only valid if $n_A \ll N$. Analogous to the previous treatment, we obtain now eq 2.21.

$$\begin{aligned}\mu^{o}_{sol,A} &= -RT\ln\left(\frac{8\pi^2}{C^{o}\sigma_A\sigma_S^{n_A}}\prod_{i=1}^{M_A+Mn_A}(2\pi m_i RT)^{3/2}\int e^{-(U(\mathbf{r}_A,\mathbf{r}_{AS})+\mathcal{W}(\mathbf{r}_A,\mathbf{r}_{AS}))/RT}\mathrm{d}\mathbf{r}_A\mathrm{d}\mathbf{r}_{AS}\right) \\ &\quad + p^{o}(\overline{V}_A + n_A\overline{V}_S) - n_A\mu_S \\ &= \mu^{o}_{n_A,A} - n_A\mu_S\end{aligned} \tag{2.21}$$

Here, $\mathbf{r}_A$ and $\mathbf{r}_{AS}$ represent respectively all coordinates of the solute A and of the n_A solvent molecules associated with A; M_{n_A} is the number of atoms of the n_A solvent molecules represented explicitly, M_A is the number of atoms of the solute A. The last line of eq 2.21 defines the standard chemical potential μ_{n_A} of the solute molecule with n_A explicit solvent molecules. This equation shows that, although an implicit solvent representation is used, a few explicit water molecules can also be included into the calculation. Only the cost of removing the solvent molecules from the bulk solvent is needed to correct the chemical potential.

The standard free energy of binding is then defined as

$$\Delta G^{o} = \mu_{AB,n_{AB}} - \mu_{A,n_A} - \mu_{B,n_B} - (n_{AB} - n_A - n_B)\mu_S \tag{2.22}$$

where n_{AB}, n_A, and n_B are the numbers solvent molecules associated explicitly with AB, A, and B, respectively. Actually the water molecules represented explicitly need to sample the whole available configurational space. If constrains are applied, the artificial effects of these constrains need to be corrected too.

2.3 Electrostatic Potential of Proteins in Ionic Solutions

One example for implicit solvent models is the electrostatic model that is based on Poisson-Boltzmann equation. It relies on the assumption, that a protein can be modeled as a low dielectric cavity with fixed charges in a high dielectric environment with mobile charges. The

bases of this model are continuum electrostatics and statistical mechanics (Jackson, 1975; Hill, 1960). This electrostatic model has been successfully applied to calculate energetic quantities of proteins and other molecules in solution (Sharp & Honig, 1990; Yang & Honig, 1992; Sharp & Honig, 1995; Honig & Nicholls, 1995).

The electrostatic potential $\phi(\mathbf{r})$ of the charge distribution $\rho(\mathbf{r})$ in vacuum at a position $\mathbf{r}$ is given by eq 2.23.

$$\phi(\mathbf{r}) = \int \frac{\rho(\mathbf{r})}{4\pi\varepsilon_o \mathbf{r}} d\mathbf{r} \tag{2.23}$$

where ε_o is the vacuum dielectric constant. The electrostatic field $\mathbf{E}$ is the negative gradient of the potential ϕ (eq 2.24).

$$\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r}); \qquad E_x(\mathbf{r}) = -\frac{\partial\phi(\mathbf{r})}{\partial x},\ E_y(\mathbf{r}) = -\frac{\partial\phi(\mathbf{r})}{\partial y},\ E_z(\mathbf{r}) = -\frac{\partial\phi(\mathbf{r})}{\partial z} \tag{2.24}$$

In a dielectric medium, the electrostatic field is shielded due to the reorientation of permanent and induced dipoles. It is useful to define a dielectric displacement vector $\mathbf{D}$ (eq 2.25).

$$\mathbf{D}(\mathbf{r}) = \varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) \tag{2.25}$$

where $\varepsilon(\mathbf{r})$ is the spatially varying dielectric constant. If the Gauß law is applied to a charge density $\rho(\mathbf{r})$ in a medium with the dielectric constant $\varepsilon(\mathbf{r})$, one obtains

$$\nabla\mathbf{D}(\mathbf{r}) = \nabla[\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})] = 4\pi\rho(\mathbf{r}) \tag{2.26}$$

With eq 2.24, one obtains the Poisson equation for a medium with a spatial varying dielectric constant $\varepsilon(\mathbf{r})$.

$$\nabla[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \tag{2.27}$$

A protein is assumed to have a definite structure. The charge density $\rho(\mathbf{r})$ of the solution consists therefore of two parts, the fixed charges $\rho_f(\mathbf{r})$ of the protein and the mobile charges $\rho_m(\mathbf{r})$ of the solution (ions) (eq 2.28)

$$\rho(\mathbf{r}) = \rho_f(\mathbf{r}) + \rho_m(\mathbf{r}) \tag{2.28}$$

The mobile charges of the solution are distributed according to Boltzmann statistics. The mean concentration $\langle c_i(\mathbf{r})\rangle$ of ions of type i at the position $\mathbf{r}$ is given by eq 2.29

$$\langle c_i(\mathbf{r})\rangle = c_i^{\text{bulk}} \exp(-\mathcal{W}_i(\mathbf{r})/RT) \tag{2.29}$$

where c_i^{bulk} is the bulk concentration of the ionic species i and $\mathcal{W}_i(\mathbf{r})$ is the potential of mean force experienced by an ion of type i at position $\mathbf{r}$. For a dilute solution, the potential of mean force is given by eq 2.30.

$$\mathcal{W}_i(\mathbf{r}) = Z_i e_o \phi(\mathbf{r}) \tag{2.30}$$

where Z_i is the charge number of the ion of type i and e_o is the elementary charge. The charge density $\rho_m(\mathbf{r})$ is then given by a sum over the charge density of all K ionic species in the solution (eq 2.31).

$$\rho_m(\mathbf{r}) = \sum_{i=1}^{K} c_i^{\text{bulk}} Z_i e_o \exp\left(\frac{-Z_i e_o \phi(\mathbf{r})}{RT}\right) \tag{2.31}$$

The Poisson-Boltzmann equation for a protein in ionic solution is then given by combining eqs 2.27, 2.28, and 2.31 to eq 2.32

$$\nabla[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\left(\rho_f(\mathbf{r}) + \sum_{i=1}^{K} c_i^{\text{bulk}} Z_i e_o \exp\left(\frac{-Z_i e_o \phi(\mathbf{r})}{RT}\right)\right) \tag{2.32}$$

For small electrostatic potentials $(\phi(\mathbf{r})/RT < 1)$, the Poisson-Boltzmann equation can be linearized by expanding the exponential up to the linear term (eq 2.33).

$$\sum_{i=1}^{K} c_i^{\text{bulk}} Z_i e_o \exp\left(\frac{-Z_i e_o \phi(\mathbf{r})}{RT}\right) \cong \sum_{i=1}^{K} c_i^{\text{bulk}} Z_i e_o - \sum_{i=1}^{K} c_i^{\text{bulk}} Z_i^2 e_o^2 \frac{\phi(\mathbf{r})}{RT} \tag{2.33}$$

The first term in eq 2.33 vanishes because of electroneutrality of the solution. It is useful to define the ionic strength I (eq 2.34) and the inverse Debye length κ (eq 2.35).

$$I = \frac{1}{2}\sum_{i=1}^{K} c_i^{\text{bulk}} Z_i^2 \tag{2.34}$$

$$\kappa = \sqrt{\frac{4\pi e_o^2 I}{\varepsilon(\mathrm{r})RT}} \tag{2.35}$$

With the definitions in eqs 2.34 and 2.35, and with eqs 2.32 and 2.33, one obtains the linearized Poisson-Boltzmann equation (eq 2.36).

$$\nabla[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho_f(\mathbf{r}) + \varepsilon(\mathrm{r})\kappa^2\phi(\mathbf{r}) \tag{2.36}$$

Analytical solutions of the Poisson-Boltzmann equation exist only for simple geometries (Kirkwood, 1934; Daune, 1997). For irregular geometries, solutions can be obtained by numerical methods. Most often, the Poisson-Boltzmann equation is solved by finite difference methods (see Honig & Nicholls, 1995 for review). But also other methods to solve partial differential equations such as boundary element methods (Zauhar & Varnek, 1996) or multigrid-based methods (Holst *et al.*, 1994) can be used to solve the Poisson-Boltzmann equation.

2.4 Simulation of Protein-Protein Docking

Together with Ernst-Walter Knapp and Nenad M. Kostić, I investigated the association of plastocyanin and cytochrome f. This work was a docking study in three stages. In the first stage, 32,000 trajectories, in which plastocyanin approached cytochrome f, were generated. The molecular configurations having relatively low energies were clustered into six families by structural similarity. In the second stage, six configurations having the lowest energies, one from each family, were used as starting point in a molecular dynamics simulation, for 260 ps. The whole plastocyanin molecule and the relevant parts of the cytochrome f molecule kept mobile. Water was treated explicitly. In the third stage, the following three contributions to the energy of binding were calculated: polarization of water by the charges of the proteins, determined from numerical solutions of the Poisson-Boltzmann equation; non-electrostatic (van der Waals and other non-bonded) interactions involving the proteins and water; and the Coulomb

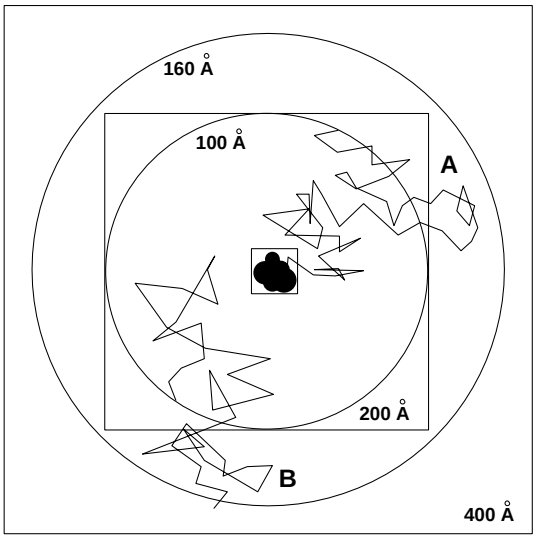

Figure 2.1: The Monte Carlo procedure for finding docking configurations. The energy of plastocyanin in a Coulomb potential of cytochrome *f* (at the center) was calculated using a finer grid within the inner cube (edge length 200 Å) and on a coarser grid elsewhere in the outer cube (edge length 400 Å). Trajectories started randomly on the inner sphere (having the radius of 100 Å). For example, trajectory A ended unproductively when plastocyanin exited the outer sphere (having the radius of 160 Å), and trajectory B ended when plastocyanin found a local minimum of energy near cytochrome *f* without entering the excluded volume.

interactions within and between the protein molecules. The total energy of association was calculated with a thermodynamic cycle. Several sets of values for ionic strength and dielectric constants gave consistent results.

The crystal structure of the luminal domain of turnip cytochrome *f* (Martinez *et al.*, 1994; PDB entry 1ctm) and structure of bean plastocyanin (Moore *et al.*, 1991; PDB entry 9pcy) were used. The two C-terminal residues absent from the structure of cytochrome *f* were modeled by a simulated annealing procedure based on molecular dynamics, while the rest of the protein was kept rigid.

Charges of most atoms, including those of the heme (Mishra *et al.*, 1983), were taken from the parameters of the CHARMM19 energy function (Brooks *et al.*, 1983). Carbon and hydrogen atoms bonded to each other were united. Charges of the copper atom and of its ligands, calculated by a density functional method, were kindly supplied by Professor E. I. Solomon (Ullmann *et al.*, 1997b). Charges of titratable groups are those at pH 7.0, assuming standard pK_a values. Electrostatic calculations by a published method (Bashford, 1991; Beroza *et al.*, 1991) showed His142 in cytochrome *f* to be electroneutral at pH 7.0.

2.4.1 Monte Carlo Docking

In the search for the docking orientation, the protein molecules were treated as rigid bodies. Center of mass of cytochrome *f* was placed at the center of two concentric cubes having edges of 200 and 400 Å and of two concentric spheres having radii of 100 and 160 Å. Dielectric constant was 80.0 everywhere. The grid in the small box encompassing cytochrome *f*, which was used in testing whether plastocyanin entered the excluded volume, had the spacing of 0.5 Å. The

Coulomb potential of cytochrome *f* was mapped on two cubic grids, with spacings of 1.0 Å in the inner box and of 2.0 Å in the outer, each having 200 x 200 x 200 points (see Figure 2.1). The energy of plastocyanin in this electrostatic potential was evaluated by multiplying the atomic charges with the potential values, obtained by linear interpolations at the atomic positions.

The configurations of the diprotein complex were simulated by an annealing procedure. Initial positions and orientations of plastocyanin were chosen randomly, by placing its center of mass on the surface of the inner sphere. Each step of the trajectory consisted of a random displacement and a random rotation. The components of these two vectors were taken from normal distributions with means of zero and with respective deviations of 2.0 Å and 5.0°. Each step was accepted or rejected according to the Metropolis algorithm (Metropolis *et al.*, 1953). If the energy of the new configuration was lower than that of the preceding one, the step was always accepted. If the new energy was higher by ΔE than the preceding one, the probability to accept the new configuration was $\exp(-\Delta E/RT)$. If plastocyanin entered the excluded volume of cytochrome *f*, the step was rejected. A total of 32,000 trajectories were created. After 3,000 accepted steps at 300 K, the temperature was lowered by 10 K. After each 300 accepted steps thereafter the temperature was again lowered by 10 K. A trajectory was unproductive and terminated if the distance between the centers of mass of the two proteins exceeded 160 Å (that is, if plastocyanin left the outer sphere). A productive trajectory was ended if in 5,000 consecutive steps the new configuration was rejected; that was the criterion for a local minimum. At 0 K the steps were continued until either of these criteria was met. The procedure just described yielded approximately 5,000 different diprotein configurations. The Coulomb energies of all those protein configurations were calculated to find the one with the highest and the lowest value. One hundred and forty configurations had electrostatic energies lower than the average value of these two and were analyzed further. Those whose structures had root-mean-square (rms) deviations of 8.0 Å or less from one another were grouped together (see Appendix). Six families of configurations emerged. The configuration having the lowest Coulomb energy in each family was further refined as described below.

2.4.2 Molecular Dynamics Simulation

At this stage conformational flexibility was imparted to each of the six electrostatically-favorable configurations, and hydration was treated by explicitly water molecules. The simulations were done with the program CHARMM (Brooks *et al.*, 1983). A sphere with a diameter of 30 Å was placed at the geometric center of plastocyanin and filled with 3587 water molecules; those that overlapped with either protein were removed (see Figure 2.2). The water model TIP3P was used (Jorgensen *et al.*, 1983). Since the structure of plastocyanin had been determined by NMR spectroscopy, no water molecules were available in this structure. A total of 114 water molecules found in the crystal structure of cytochrome *f* were included; those that overlapped with atoms of plastocyanin were removed. Water was kept in the sphere by a suitable containment potential (Brooks & Karplus, 1983). Plastocyanin and the part of cytochrome *f* inside the sphere were completely flexible, subject only to the potential defined by the force field. Each atom of cytochrome *f* outside the sphere was gently fixed with a harmonic potential defined by the force constant of 0.05 $\text{kcal}\times\text{mol}^{-1}\times\text{Å}^{-1}$. The Coulomb and Lennard-Jones interactions of each atom were calculated with group switch boundary conditions. At distances less than 9.0 Å the full potential is applied, from 9.0 to 11.0 Å the potential gradually disappeared, and at distances greater than 11.0 Å atom pair interactions vanished. The water molecules within a 2.5-Å layer below the sphere surface were coupled to a heat bath at 300 K; the friction coeffi-

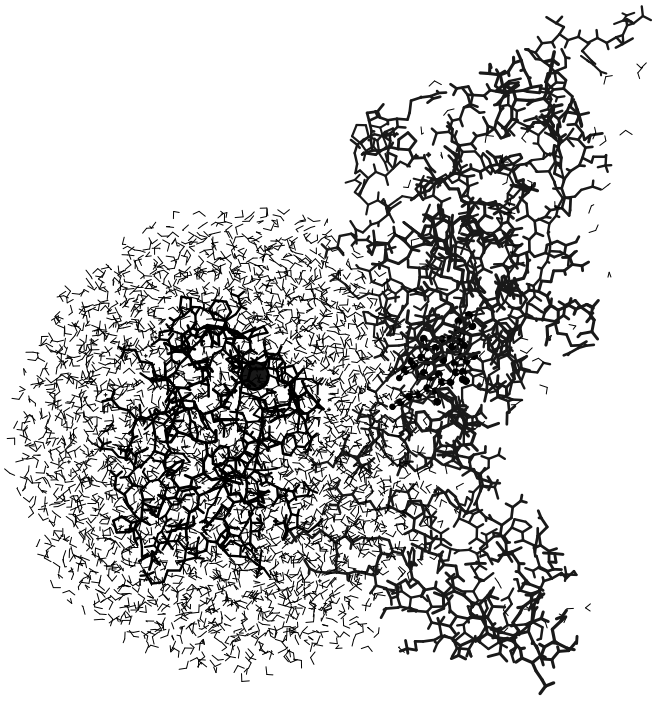

Figure 2.2: One out of six configurations of the diprotein complex between plastocyanin (left) and cytochrome *f* (right) that emerged from the Monte Carlo procedure. A sphere with the radius of 15 Å was filled with water, and molecular dynamics simulation was performed.

cient was 8 ps^{-1}. The lengths of the covalent bonds involving hydrogen atoms were constrained with the SHAKE algorithm (Ryckaert *et al.*, 1977) so that the time propagation could be done in 2-fs steps. Each of the six configurations was simulated for 200 ps at 300 K and then cooled down by coupling to heat baths at 200, 100, and 0 K. The simulations continued for 20 ps at each of these lower temperatures. Finally, each configuration was energy minimized. Each of the resulting structures was then quantitatively analyzed, as described below.

2.4.3 Energetics of Docking

At least four factors contribute to the change in free energy of the protein pair upon docking, ΔG_D, according to eq 2.37.

$$\Delta G_D = \Delta\Delta G_R + \Delta G_{NE} + \Delta G_C + \Delta G_{TRC} \tag{2.37}$$

Changes in solvation energy come from the electrostatic, $\Delta\Delta G_R$,[1] and the non-electrostatic, ΔG_{NE} , interactions. There are also changes in the Coulomb electrostatic energy, ΔG_C. Finally, loss of translational, rotational, and conformational mobility is represented by ΔG_{TRC}. Each of these four terms is the energy difference between the docked complex of plastocyanin and cytochrome *f* on the one side and the free plastocyanin and cytochrome *f* on the other. The last term in eq 2.37 is likely to have similar values for all the configurations of the docked complex. Since we are interested in differences among the configurations, i. e., in their relative stabilities,

[1]The electrostatic solvation energy ΔG_R by itself is an energy difference between the energy of the unsolvated and the solvated molecule.

neglecting this term is justifiable. The remaining three terms can be estimated with the help of various approximations.

The electrostatic contribution to the solvation energy is caused by polarization of the medium by the charges in the protein. This contribution can be calculated by solving the Poisson-Boltzmann equation (Warwicker & Watson, 1982; Klapper *et al.*, 1986; Gilson *et al.*, 1987; Gilson & Honig, 1988; Honig & Nicholls, 1995; Sharp & Honig, 1990). The energy required to bring a molecule with the dielectric constant ε_s from a medium with the constant ε_m to a medium with the constant ε_s is given in eq 2.38.

$$\Delta G_R = G_R(\varepsilon_s, \varepsilon_s) - G_R(\varepsilon_s, \varepsilon_m) \tag{2.38}$$

The energy of the molecule within its reaction field is defined in eq 2.39,

$$G_R(\varepsilon_s, \varepsilon_i) = \frac{1}{2} \sum_i q_i \phi_i(\varepsilon_s, \varepsilon_i) \tag{2.39}$$

in which q_i are the atomic charges in the molecule, ϕ_i is the electrostatic potential at the position of the charge *i* inside a dielectric cavity with a dielectric constant ε_s in a medium with a constant ε_i such that ε_i is ε_m or ε_s. The boundaries of the cavity are defined by the water-accessible surface of the molecule. The potentials ϕ_i can be obtained by solving numerically the Poisson-Boltzmann equation or its linearized form.

Non-electrostatic contributions to the solvation energy are caused by van der Waals interactions between the solute and the solvent and involve also mechanical energy occurring by the creation or enlargement of the solute cavity against the solvent pressure. Therefore, the second term in eq 2.37, can be estimated by the empirical rule in eq 2.40.

$$\Delta G_{NE} = a + bA \tag{2.40}$$

The water-accessible surface *A* is defined by rolling a sphere with a radius of 1.4 Å over the protein molecule. The parameters *a* and *b* were empirically fitted (Ben-Naim, 1994; Sitkoff *et al.*, 1994). Since we are interested in relative, not absolute, energies of various docking configurations, the parameter *a*, which is the same for all of them, can be neglected. The parameter *b* was set at 5.0, 6.8, and 20.0.

Coulomb energy arises from the pairwise interactions between the atoms within each protein and between the two proteins. It can be calculated according to eq 2.41, in which *i* and *j* run over all the atoms.

$$\Delta G_C(\varepsilon_s) = \frac{1}{2} \sum_{i,j} \frac{q_i q_j}{\varepsilon_s r_{ij}} \tag{2.41}$$

The electrostatic energy (ΔG_E) is defined as the sum of the so-called reaction field and the Coulomb energy (in eq 2.42) and the total energy (ΔG_T) as the sum of all three terms, i. e., of the electrostatic and non-electrostatic contributions (in eq 2.43).

$$\Delta G_E = \Delta\Delta G_R + \Delta G_C \tag{2.42}$$

$$\begin{aligned} \Delta G_T &= \Delta\Delta G_R + \Delta G_C + \Delta G_{NE} \\ &= \Delta G_E + \Delta G_{NE} \end{aligned} \tag{2.43}$$

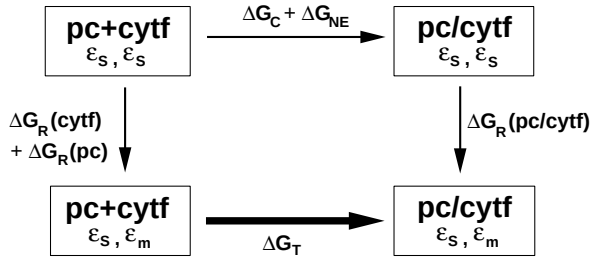

Figure 2.3: Thermodynamic cycle to calculate the relative association energy of proteins. With this type of thermodynamic cycle, also conformational changes upon binding are allowed.

The energy difference arising from the loss of translational, rotational, and conformational mobility is neglected in eq 2.43 compared to eq 2.37. In eq 2.38, 2.39, and 2.41, ε_m was set at 80.0 for water, and ε_s was set at 1.0, 2.0, and 4.0 for the protein matter. In eqs 2.38 and 2.39, the ionic strength was set at 0.10 M, and an ion exclusion layer of 2.0 Å was used.

Numerically solving the Poisson-Boltzmann equation requires to assign the charge distribution discretely to a grid. This gives rise to artifacts, termed grid energy, which preclude a direct calculation of the docking energy by considering the separate and associated proteins (see the thick horizontal line in Figure 2.3). These problems can be avoided by calculating the other three steps in the thermodynamic cycle. The energies required to transfer the separate proteins from a medium with the dielectric constant of water to a medium with the dielectric constant of the protein and the energy required to transfer each of the six optimized configurations of the diprotein complex in the reverse direction were calculated. With this procedure, the grid energies cancels, and quantities for the two vertical processes in Figure 2.3 are obtained. Because the dielectric constants of the solute and the medium are equal, the ΔG_C term was calculated with the Coulomb law, using a single dielectric constant. Estimation of ΔG_{NE} completes the cycle and allows the evaluation of ΔG_T. This procedure was repeated for each of the six configurations of the diprotein complex; the structures of the separate proteins were kept the same.

Conformations of the individual proteins, and also configuration of the diprotein complex, are flexible in this study. Other thermodynamic cycles, which are applicable to rigid structures (Jackson & Sternberg, 1995), are inapplicable here because the grid energy would not cancel exactly. The structures of the plastocyanin-cytochrome f complex obtained from the simulation and possible consequences for the electron transfer reaction are discussed in Section 5.3.

2.5 Similarity of Isofunctional Proteins

Different proteins that dock and react with the same reaction partner in a similar manner are likely to bind with a similar pattern. Three-dimensional similarity indices as the integral-based Hodgkin index H_{AB}^{elec} (eq 2.44) can be used to compare molecules and provide a quantitative measure of the similarity of two molecules (Good, 1995).

$$H_{\mathrm{AB}}^{elec} = \frac{2\int \phi_{\mathrm{A}}\phi_{\mathrm{B}}\mathrm{d}V}{\int \phi_{\mathrm{A}}{}^2\mathrm{d}V + \int \phi_{\mathrm{B}}{}^2\mathrm{d}V} \tag{2.44}$$

The Coulomb potentials ϕ of the structurally different molecules A and B are integrated over the whole volume. The numerator quantifies the spatial overlap of the two electrostatic potentials ϕ_{A} and ϕ_{B}, while the denominator normalizes this value. The resulting similarity index falls in the interval between -1 and $+1$. The value $+1$ corresponds to molecules with identical

potentials, whereas -1 corresponds to electrostatic complementarity, this means the potentials are of the same magnitude but have opposite sign. The electrostatic potential ϕ at the position $\mathbf{r}$ in a medium with the dielectric constant ε is calculated from point charges q_i assigned to each atom i as given by eq 2.45; $\mathbf{r_i}$ is the positional vector of atom i.

$$\phi(\mathbf{r}) = \sum_{i=1}^{n} \frac{q_i}{\varepsilon \mid \mathbf{r} - \mathbf{r}_i \mid} \cong \sum_{i=1}^{n} \frac{q_i}{\varepsilon} \sum_{j=1}^{k} G_i^j(\mathbf{r}) \tag{2.45}$$

$$G_i^j(\mathbf{r}) = \upsilon_j \exp(-\varsigma_j(\mathbf{r} - \mathbf{r}_i)^2) \tag{2.46}$$

The $1/r$ term of the Coulomb law can be approximated by a spherical Gaussian function (eq 2.46). This approximation avoids the singularities of the Coulomb terms and makes the computation about one hundred times faster than those involving grid evaluation (Good *et al.*, 1992). In general, two Gaussians are sufficient to fit a $1/r$ curve within a range from 2.0 to 12.0 Å. By applying standard least-squares fitting techniques with a spherical shell weighting function (Shavitt, 1962), the following parameters are obtained (Hauswald, 1998; Ullmann *et al.*, 1997a): $\varsigma_1 = 0.1247$ Å^{-2}, $\varsigma_2 = 0.0065$ Å^{-2}, $\upsilon_1 = 0.5168$ Å^{-1}, and $\upsilon_2 = 0.1958$ Å^{-1}. With these parameters, the Hodgkin index H_{AB}^{elec} accounts for a significant portion of the electrostatic potentials outside the two protein molecules. The series of integrals reduce to spherical Gaussians with modified prefactors and exponents, which depend only on the pairwise interatomic distances between the two proteins, i.e., on the relative orientation of these molecules. The parameters for these Gaussians can be easily calculated (Hauswald, 1998). The analytical formula for the Hodgkin index can be evaluated extremely rapidly and no singularities appear in this approximation. In a homogeneous medium, the Hodgkin index based on the Coulomb electrostatic potential is independent of the dielectric constant, since the dielectric constant cancels in eq 2.44.

Together with Markus Hauswald, Axel Jensen, Nenad M. Kostić, and Ernst-Walter Knapp, I applied the Hodgekin Index to align the isofunctional proteins plastocyanin and cytochrome c_6 (Ullmann *et al.*, 1997a) and the isofunctional proteins ferredoxin and flavodoxin (Ullmann *et al.*, 1998a). The method we applied is termed FAME (Flexible Alignment of Molecule Ensembles) (Hauswald, 1998). In contrast to applications in drug design, the structure of the proteins was not varied in this approach.

We started from one hundred different random initial orientations of the two proteins with coinciding centers of mass and maximized the Hodgkin index (eq 2.44) with respect to translational and rotational degrees of freedom to ensure to find the absolute maximum several times. The highest-ranked alignment can be taken to represent the global maximum of the Hodgkin index (eq 2.44). The rotations were parameterized in quaternions, as previously done in SEAL (Kearsley & Smith, 1990). Quaternions behave correctly if the rotation degenerates to the identity operator, while Euler angles are undetermined in this case (Altmann, 1986). The FAME method involves a highly efficient eigenvector-following algorithm (Baker, 1992) for optimizations, which requires analytic first and second derivatives of the function H_{AB}^{elec}. Typically, 15 steps (minimal 8, maximal 28) were needed to converge to a local maximum of this function. Convergence was reached if the norm of the gradient became less than 10^{-10}. One maximization took about 0.5 h on a MIPS R8000 CPU. Here, I have summarized only the keypoints of the method that are necessary to reproduce our results. For a detailed description of the FAME method and additional features of the program see Hauswald, 1998. The results of the superposition of plastocyanin and cytochrome c_6 are discussed in Section 5.4, those of the superposition of flavodoxin and ferredoxin in Section 5.6.

superposition method	protein	Hodgkin index[a]
FAME	c6	0.85
	pc	0.92
	az	0.56
dipole alignment	c6	0.75
Kabsch algorithm	pc	0.91

a — defined in eq 2.44

Table 2.1: Hodgkin indices for the superpositions of *Chlamydomonas reinhardtii* plastocyanin with *Chlamydomonas reinhardtii* cytochrome c_6 (c6), poplar plastocyanin (pc) and *Pseudomonas aeruginosa* azurin (az) obtained by three different alignment methods (FAME, dipole alignment, Kabsch algorithm);

The properties of plastocyanin and cytochrome c_6 or of ferredoxin and flavodoxin are expected to be most similar, if these proteins are from the same organism. Only for one species, respectively, the structures of both proteins of the two isofunctional couples are known. Plastocyanin and cytochrome c_6 are known from *Chlamydomonas reinhardtii* (Kerfeld *et al.*, 1995; Redinbo *et al.*, 1993) and ferredoxin and flavodoxin are known from *Anabaena* PCC 7120 (Rypniewski *et al.*, 1991; Rao *et al.*, 1993). We used the PDB entries 2plt for plastocyanin, 1cyj for cytochrome c_6, 1fxa for ferredoxin, 1rcf for flavodoxin. To test the applicability of the FAME algorithm to superimpose proteins, we used it to align *Chlamydomonas reinhardtii* plastocyanin with poplar plastocyanin (Guss *et al.*, 1992) (PDB entry 1plc, sequence identity with 2plt: 60 %) and with *Pseudomonas aeruginosa* azurin (PDB entry 5azu, chain A) (Nar *et al.*, 1991). For the superposition of electrostatic potentials with the FAME algorithm and for the calculation of dipole vectors, we deleted all water molecules contained in the crystal structure. They may contribute non-specifically to the overall electrostatic potential, since their position depends often on the crystalline environment. The charges of most atoms were taken from the CHARMM22 parameters (MacKerell *et al.*, 1998). Atomic charges for the blue copper site in plastocyanin, calculated by a density functional method, were kindly supplied by Professor E. I. Solomon. The same charges were assigned to the blue copper site in azurin. Charges of titratable groups are those at pH 7.0, assuming standard pK_a values.

To test the utility of the FAME algorithm for superimposing proteins, we have applied it to superimpose plastocyanin from *Chlamydomonas reinhardtii* with two other blue copper proteins; with plastocyanin from poplar and with azurin from *Pseudomonas aeruginosa*. Poplar plastocyanin and *Chlamydomonas reinhardtii* plastocyanin have sequences, that are only 60 % identical, but they have very similar electrostatic properties. Plastocyanin from poplar and plastocyanin from *Chlamydomonas reinhardtii*, respectively, have charges of -8 and -6 at pH 7 and dipole moments of 330 D and 340 D. Both, azurin and plastocyanin have β-barrel folds. Although they belong to the same structural family (Adman, 1991), their electrostatic properties are different. The overall charge of azurin at pH 7 is -1.0, and its dipole moment has a magnitude of only 95 D. In spite of their dissimilar electrostatic properties, both, azurin and plastocyanin, have a hydrophobic patch close to the copper site.

The relative orientation of the two plastocyanins found with the FAME algorithm is very similar to an orientation found with the Kabsch algorithm (Kabsch, 1976; Kabsch, 1978), which minimizes the distances between equivalent backbone atoms of two proteins. As Table 2.1

shows, the Hodgkin indices (eq 2.44) of the superpositions, found with the FAME algorithm (0.92) and with the Kabsch-algorithm (0.91) are nearly equal. The best Hodgkin index found for *Chlamydomonas reinhardtii* plastocyanin and *Pseudomonas aeruginosa* azurin (0.56) is considerable lower than that for the superposition of the two plastocyanins. Although the electrostatic potentials of plastocyanin and azurin are rather different, their hydrophobic patches overlap in the orientation with the highest Hodgkin index. This overlap shows, that analogous surface regions can be identified, even when the proteins compared have different overall electrostatics.

These results illustrate, that the Hodgkin index is a good measure of the similarity of electrostatic potentials of proteins. The FAME algorithm is an unbiased and meaningful method for aligning two proteins with similar electrostatics and comparable size. The obtained alignment can be used to compare the functionality of these proteins.