Calculation of the Redox Potential of the Protein Azurin and Some Mutants

Marieke van den Bosch,[a] Marcel Swart,[b] Jaap G. Snijders†,[c] Herman J. C. Berendsen,[d] Alan E. Mark,[d] Chris Oostenbrink,[e] Wilfred F. van Gunsteren,[e] and Gerard W. Canters*[a]

Azurin from Pseudomonas aeruginosa *is a small 128-residue, copper-containing protein. Its redox potential can be modified by mutating the protein. Free-energy calculations based on classical molecular-dynamics simulations of the protein and from mutants in aqueous solution at different pH values were used to compute relative redox potentials. The precision of the free-energy calculations with the λ coupling-parameter approach is evaluated as function of the number and sequence of λ values, the sampling time and initial conditions. It is found that the precision is critically dependent on the relaxation of hydrogen-bonding networks when changing the atomic-charge distribution due to a change of redox state or pH value. The errors in the free energies range from 1 to 10* k_BT*, depending on the type of process. Only qualitative estimates of the change in redox potential by protein mutation can be obtained.*

Introduction

Azurin from *Pseudomonas aeruginosa* (PA) is a small 128-residue, copper-containing redox protein that plays a role in electron transport.[1] In the case of redox proteins, there is considerable interest in understanding how the protein matrix acts to tune the redox potential of the metal site. Three-dimensional structures for azurin PA have been determined for different pH values.[2–4] Apart from a small conformational change at residues 35 to 37, which is due to (de)protonation of His35, the protein structures at different pH values are rather similar (Figure 1). The copper ion is coordinated by the S_γ atom of Cys112, the N_δ atoms of His46 and His117, the carbonyl oxygen of Gly45 and the S_δ atom of Met121; this results in a trigonal bipyramidal coordination geometry around the copper. Structures of the reduced copper site[3,6,7] are nearly identical to those of the oxidised protein. Experimentally, azurin has been extensively studied, and the effects of a large number of mutations on the redox potential are known.[8–10] To complement such studies there is much interest in computational methods that could be used to estimate changes in redox potential associated with changes in pH or specific mutations. To directly calculate the redox potential of azurin as a function of pH or protein composition is, however, currently not possible. This would require a full quantum-chemical treatment of the copper site, including its immediate environment, and would need to include entropic effects associated with alternative conformational states of the protein and the solvent. Since extensive sampling of atomic motion in combination with high-level electronic-structure calculations is not practicable for systems of the required size and complexity, one must make one or more simplifying assumptions:

i) One may treat the copper ion and its ligands using high-level quantum-chemical methodology, but neglect the protein environment, the solution and entropic contributions.[11–13]

ii) One may treat the protein in explicit solvent using an empirical force field and classical dynamics. This approach includes entropic effects through extensive conformational sampling. It treats the processes of oxidation and reduction as purely electrostatic effects, implicitly assuming that quantum-mechanical contributions are the same for all mutations.[14–18]

iii) A third approach is to simply neglect both the quantum-mechanical nature of the transition and entropic contributions from the environment and treat oxidation and reduction as purely electrostatic phenomena. For example, one may perform Poisson–Boltzmann calculations based on a given (X-ray) structure in the oxidised and reduced states.[19,20]

A combination of different approaches might be the most appropriate method.[21–25] However, combined approaches introduce inconsistencies between different aspects of the prob-

[a] *M. van den Bosch, Prof. G. W. Canters*
Leiden Institute of Chemistry, Gorlaeus Laboratories
Einsteinweg 55, 2333 CC Leiden (The Netherlands)
Fax: (+31) 71-527-4349
E-mail: canters@chem.leidenuniv.nl

[b] *Dr. M. Swart*
Department of Organic and Inorganic Chemistry, Free University
de Boelelaan 1083, 1081 HV Amsterdam (The Netherlands)

[c] *Prof. J. G. Snijders†*
Department of Theoretical Chemistry, Groningen University
Nijenbogh 4, 9747 AG Groningen (The Netherlands)

[d] *Prof. H. J. C. Berendsen, Prof. A. E. Mark*
Groningen Biomolecular Sciences and Biotechnology Institute
Department of Biopysical Chemistry, University of Groningen
Nijenbogh 4, 9747 AG Groningen (The Netherlands)

[e] *Dr. C. Oostenbrink, Prof. W. F. van Gunsteren*
Laboratory of Physical Chemistry
Swiss Federal Institute of Technology, ETH Hönggerberg
8093 Zürich (Switzerland)

  DOI: 10.1002/cbic.200400244 *ChemBioChem* **2005**, *6*, 738 – 746

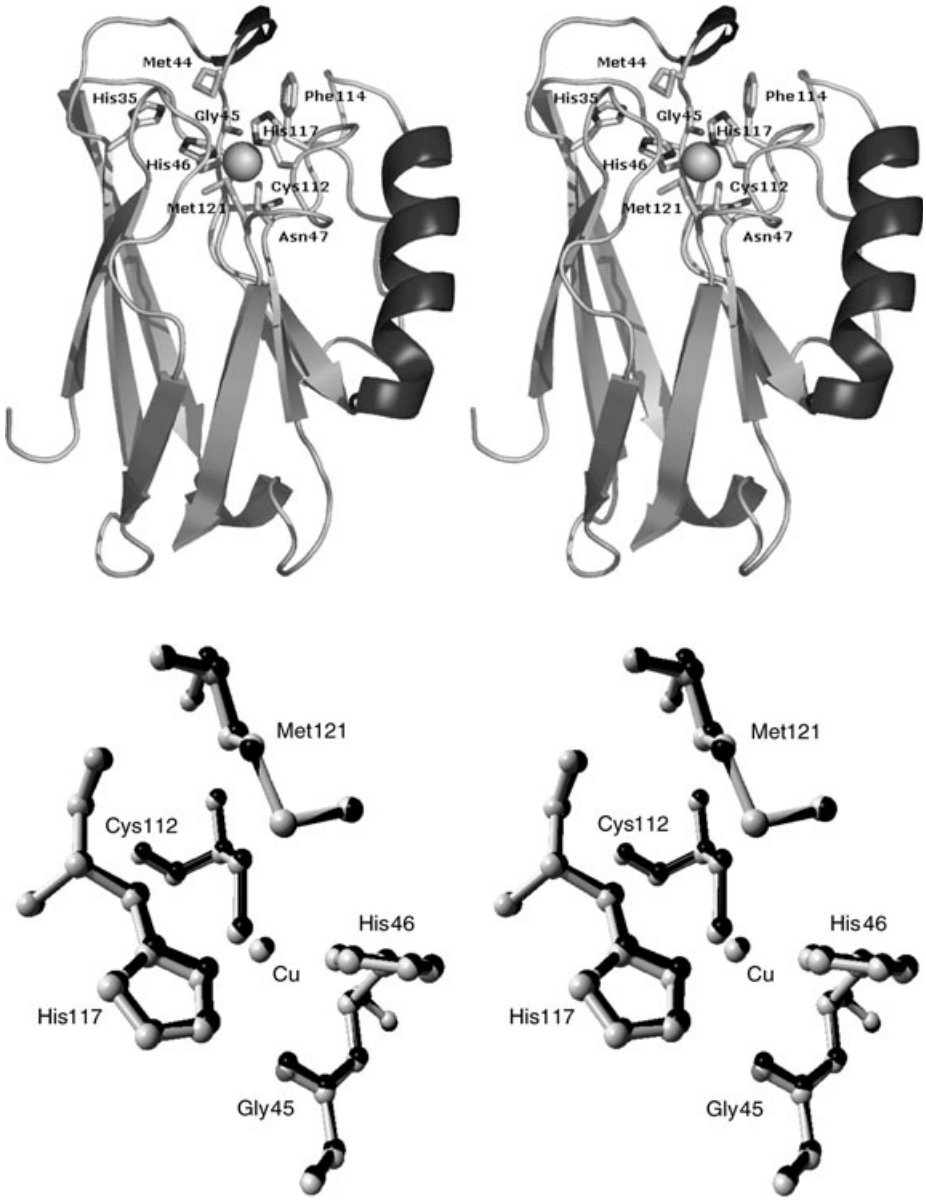

Figure 1. *Top) X-ray structure[3] in stereo presentation (PDB[5] code 4AZU, subunit A) of azurin in its oxidised form at pH 5. The Cu atom, its ligands and the side chains that are mutated are represented at the atomic level. Bottom) Overlay of the active sites of the oxidised (black) and reduced (grey) forms in ball and stick mode.*

lem and make it impossible to determine the dominant contributions. In this article, we explore the accuracy and applicability of the second approach. Molecular-dynamics (MD) simulations of azurin in both oxidised and reduced forms and at pH 5 and pH 9 in explicit water are used to estimate the relative redox potential of the wild-type protein and four mutations: M44K, N47L, N47D and F114A.

The thermodynamic integration method[26] was used to compute free-energy differences between different states of the system.[27] In this approach, the Hamiltonian H(**p**,**r**) that describes the molecular system as function of the generalised coordinates **r** and momenta **p** is made a function of a coupling parameter λ, H(**p**,**r**;λ). The coupling parameter is chosen such that when $\lambda = \lambda_A$, the Hamiltonian of the molecular system corresponds to that of state A, that is, $H(\mathbf{p},\mathbf{r};\lambda_A) = H_A(\mathbf{p},\mathbf{r})$ and when $\lambda = \lambda_B$, the Hamiltonian of the system corresponds to that of state B, that is, $H(\mathbf{p},\mathbf{r};\lambda_B) = H_B(\mathbf{p},\mathbf{r})$. Since the Hamiltonian is a function of λ, the free enthalpy or Gibbs free energy[27]

$$G(N,p,T) = -k_B T \ln[h^{-3N} V^{-1} \iiint \exp\{-(H(\mathbf{p},\mathbf{r}) + pV)/k_B T\} d\mathbf{p} d\mathbf{r} dV] \quad (1)$$

is also a function of λ. The derivative of the free energy with respect to λ is given by[26]

$$dG(\lambda)/d\lambda = \langle \partial H(\lambda)/\partial\lambda \rangle_\lambda \quad (2)$$

Here N is the number of particles in the system, p is the pressure and V is the volume of the system, T is the temperature, k_B is Boltzmann's constant and h is Planck's constant. From Equation (2), the thermodynamic-integration (TI) formula for the free-energy difference between states A and B follows immediately

$$\Delta G = G(\lambda_B) - G(\lambda_A) = \int_{\lambda A}^{\lambda B} \langle \partial \mathrm{H}(\lambda)/\partial\lambda \rangle_\lambda \mathrm{d}\lambda \qquad (3)$$

To evaluate Equation (3), the ensemble average $\langle \partial H/\partial\lambda \rangle$ can be determined from simulations at a series of λ values between λ_A and λ_B, and the integral can be determined numerically. As λ can refer to any arbitrary coordinate, the coupling-parameter approach may be used to link two physical states through a nonphysical pathway. Here, the charge distribution on the copper and its ligands is varied as function of λ to model the transition from the oxidised to the reduced state of azurin. Alternatively, the protonation state of the residue His35 was altered as a function of λ to mimic the transition from pH 5 to pH 9. The protonation state of His83 was not varied, since the side chain of this residue is exposed to the solvent. The functional dependence of H(**p**;**r**;λ) on λ effectively describes the pathway from state A to state B.

The computation of a free-energy difference by using Equation (3) involves three basic choices:

i) the definition of the end states A and B,
ii) the pathway that will connect them, and
iii) the definition of one or more thermodynamic cycles of which they are a part.

Since the free energy is a thermodynamic state function, the change in free energy $\Delta G_{B,A}$ will be independent of the path connecting states A and B. Along any closed path or cycle, $\Delta G_{cycle} = 0$. This is true so long as the system remains in equilibrium, and the change is performed in a reversible manner. Thus, if we consider the cycle (4) and can calculate all four legs it, we may use cycle closure to test the quality of the equilibration, sampling and integration over λ.

azurin (reduced, pH 5) ⟷ azurin (oxidised, pH 5)
⇕ ⇕ (4)
azurin (reduced, pH 9) ⟷ azurin (oxidised, pH 9)

The study presented in this paper consists of four parts. First, the stability of the protein in the four (end) states of cycle (4) is investigated. Second, the convergence behaviour of $\langle \partial H/\partial\lambda \rangle$ as a function of the simulated time per λ value and as a function of the choice of initial atomic velocities is examined. We investigate the precision of expression (3) as function of the number of discrete λ values used and of the choice of the sequence of the particular λ values when changing the system from one λ value to the next. The thermodynamic cycle (4) was used to estimate the precision of the calculated ΔG values. Third, the effects of variations in the Hamiltonian on the calculated ΔG values are considered. Fourth, the thermodynamic-integration technique is applied to investigate the effect of various azurin mutations on the redox potential.

Results and Discussion

Overall structure of the protein

To test the stability of the protein within the force field, the four end states of thermodynamic cycle (4) were equilibrated for 450 ps then simulated for 400 ps. The protein structures in all cases remained close to the X-ray structure of the oxidised form at pH 5[3] with backbone atom-positional root-mean-square differences after 950 ps with respect to this X-ray structure of 0.12 nm for ox, pH 5; 0.11 nm for red, pH 5; 0.17 nm for ox, pH 9 and 0.14 nm for red, pH 9. The secondary structure as derived from NMR experiments for the reduced form at pH 5.5 and 40 °C in aqueous solution[4] is also largely maintained in the four simulations of the four end states. This is illustrated in Table 1 in which the percentage backbone–backbone NH···CO

Table 1. *Populations [%] of main-chain hydrogen bonds in MD simulations of azurin in water in different redox states and at different pH values.*[a]

NH–CO averaging period [ps]	red, pH 5 400/400	ox, pH 5 400/400	red, pH 9 400/400	ox, pH 9 800/800
Turn type II				
4–1	0/0	0/0	0/0	0/0
strands S1–S3				
30–3	98/99	98/97	96/87	97/99
5–30	97/96	97/100	98/93	99/99
32–5	90/97	94/98	99/98	99/99
7–32	96/98	95/96	98/98	99/94
34–7	98/78	95/78	94/95	91/94
9–34	95/84	98/93	94/1	86/99
turn type I				
13–10	0/0	0/0	0/0	0/0
strands S2a–S1				
10–14	28/82	6/71	60/57	1/77
16–8	94/57	89/98	81/98	98/66
turn type II				
19–16	0/0	0/0	0/0	0/0
strands S2b–S8				
124–18	89/83	92/94	91/97	92/89
20–124	82/93	84/92	90/99	95/94
126–20	91/90	87/72	74/98	76/97
22–126	98/80	98/97	98/99	97/98
128–22	96/0	92/36	96/60	91/80
24–128	93/0	91/1	80/0	34/21
turn type I				
26–23	77/7	72/91	65/14	79/88
turn type II				
28–25	0/0	0/0	0/0	0/0
strands S3–S6				
99–27	96/97	87/99	87/6	91/92
29–97	97/98	98/99	98/90	98/100
97–29	99/99	96/99	99/94	99/99
31–95	99/98	99/99	99/90	99/99
95–31	97/98	90/98	98/95	97/99
33–93	91/92	88/93	88/98	96/93
93–33	99/54	99/99	97/96	97/98

Table 1. (Continued)

NH–CO averaging period [ps]	red, pH 5 400/400	ox, pH 5 400/400	red, pH 9 400/400	ox, pH 9 800/800
35–91	75/85	97/98	99/78	98/97
helix H1				
43–40	64/68	57/62	67/54	54/60
44–40	95/98	89/99	97/97	98/96
45–41	0/0	0/0	0/0	0/0
strands S4–S5				
46–87	98/82	100/93	99/94	99/99
87–46	0/0	0/0	0/0	0/0
84–48	86/1	92/96	80/81	96/92
50–82	99/0	98/45	98/98	98/98
82–50	54/0	53/0	50/38	59/48
strands S4–S7				
49–111	96/98	97/98	94/96	97/98
111–49	97/100	98/100	96/99	98/98
51–109	97/94	99/97	98/99	97/99
109–51	94/95	88/10	94/88	81/63
53–107	76/96	52/0	40/0	17/0
helix H2				
55–52	14/49	12/3	14/0	82/60
56–53	0/0	0/0	0/0	0/0
57–54	0/0	0/0	0/0	0/0
59–55	92/0	86/0	86/0	15/0
60–56	95/86	95/83	95/90	96/93
61–57	93/94	82/89	94/97	94/86
62–58	86/72	81/66	86/94	80/82
63–59	95/99	98/0	98/0	99/99
64–60	99/99	99/0	100/0	100/100
65–61	76/90	79/0	90/0	93/93
66–63	53/37	71/0	33/24	21/15
67–64	8/40	14/0	23/0	35/32
turn type I				
71–68	50/56	44/8	34/51	47/52
73–70	89/99	83/91	95/94	88/97
turn type II				
77–74	0/0	0/0	0/0	0/0
turn type I				
80–77	7/0	25/2	29/2	25/13
turn type II				
91–88	77/0	48/58	76/86	76/75
helix H3				
101–98	40/4	6/0	59/3	12/12
102–99	77/89	71/87	84/2	79/82
turn type II				
106–103	17/39	31/29	34/0	75/33
strands S7–S8				
108–125	17/85	47/87	62/89	83/96
125–108	98/96	99/100	97/98	99/99
110–123	99/98	99/97	99/99	99/99
123–110	98/99	96/100	99/97	99/97
112–121	13/35	38/54	38/58	77/17
turn type II				
117–114	73/70	60/44	72/72	44/43
helix H4				
120–117	66/33	17/26	51/55	46/32
121–117	95/96	93/97	96/98	98/94

[a] The two percentages for the first three states were obtained from simulations before and after a forward and backward change to another state. For the first three columns these intermediate states are: ox, pH 5; ox, pH 9 and ox, pH 9. The percentages of the ox, pH 9 state are from simulations after a change from the ox, pH 5 state and from the red, pH 9 state. A hydrogen bond is defined by a hydrogen–acceptor distance of less than 0.25 nm and a donor-hydrogen-acceptor angle larger than 135°. Only hydrogen bonds indicated in Figure 6 of ref. [4] have been listed.

hydrogen bonding in the simulations is presented for the hydrogen bonds that were inferred from the NMR experiment (Figure 6 of ref. [4]). Only nine out of 68 backbone–backbone hydrogen bonds were not seen in the simulations. Five are located in turns, which may not have a very stable conformation. Two are i to $i-3$ hydrogen bonds (i is the residue sequence number) in helix 2 and one is an i to $i-4$ hydrogen bond at the N terminus of the short helix 1. Of the many interstrand hydrogen bonds, only one, from residue 87 to residue 46 at the edge of strands S4 and S5, is not seen in the simulations. Instead, the NH of residue 87 makes a stable hydrogen bond to the carbonyl oxygen of residue 47.

Free-energy calculations

In Figure 2 the values of $\langle \partial H/\partial\lambda \rangle$ along the four legs of the thermodynamic cycle (4) are shown. The ΔG values as calculated from Equation (3) are shown at the arrows along the four legs of the cycles in Figure 3. When more than one value is given, the upper value, ΔG_f, is from the forward change of λ ($\lambda = 0$ to 1 in the direction of the arrow), the middle value, ΔG_b, is from the reverse change of λ ($\lambda = 1$ to 0), and the lower value is the Boltzmann weighted average, calculated from:[28]

$$\Delta G_{ave} = -k_B T \ln \frac{\exp^{-\Delta G_f/k_B T} + \exp^{-\Delta G_b/k_B T}}{2} \quad (5)$$

with $k_B T = 2.5\ kJ\,mol^{-1}$. For the process of oxidising the copper at pH 5 (Figure 2A), most of the difference of $10\ kJ\,mol^{-1}$ between the forward and reverse processes (Figure 3A) occurs for $\lambda < 0.4$. At pH 9, the hysteresis of this process is small ($< 1\ kJ\,mol^{-1}$; Figure 2B). The curves for the process of changing the pH display a much larger hysteresis of $45\ kJ\,mol^{-1}$ in the reduced state (Figure 2C) and $30\ kJ\,mol^{-1}$ in the oxidised state. Apparently, the protein does not have sufficient time during the simulations to relax and reach equilibrium as λ is changed. This is confirmed by the results obtained when λ was varied in a nonsequential (random) manner. The green curves in Figure 2 show larger differences in $\langle \partial H/\partial\lambda \rangle_\lambda$ values for adjacent λ values than when λ was changed sequentially (red and black curves), despite the fact that values at each λ should be independent. Changing λ nonsequentially results in values for ΔG_{cycle} that are twice as large (Figure 3). Extending the number of λ values from 11 (Figure 3A and C) to 21 (B and D) in the integral of Equation (3) does not significantly improve cycle closure (ΔG_{cycle}). This is not surprising considering that the $dG/d\lambda$ profile is smooth. Extending the sampling time per λ value does not dramatically improve the convergence of $\langle \partial H/\partial\lambda \rangle_\lambda$, as is illustrated in Figure 2B. In view of this, neither is a reverse cumulative average (RCA) analysis of the data expected to provide further insight into the convergence question.[29] In Figure 2B, it is indicated that neither does varying the initial velocities per λ value, which can be used to randomise sampling, improve the overall convergence of ΔG. We have tried to identify structural changes or lack of them that are correlated with the occurrence of hysteresis at particular λ values. When considering the change from pH 5 to pH 9, for which process the

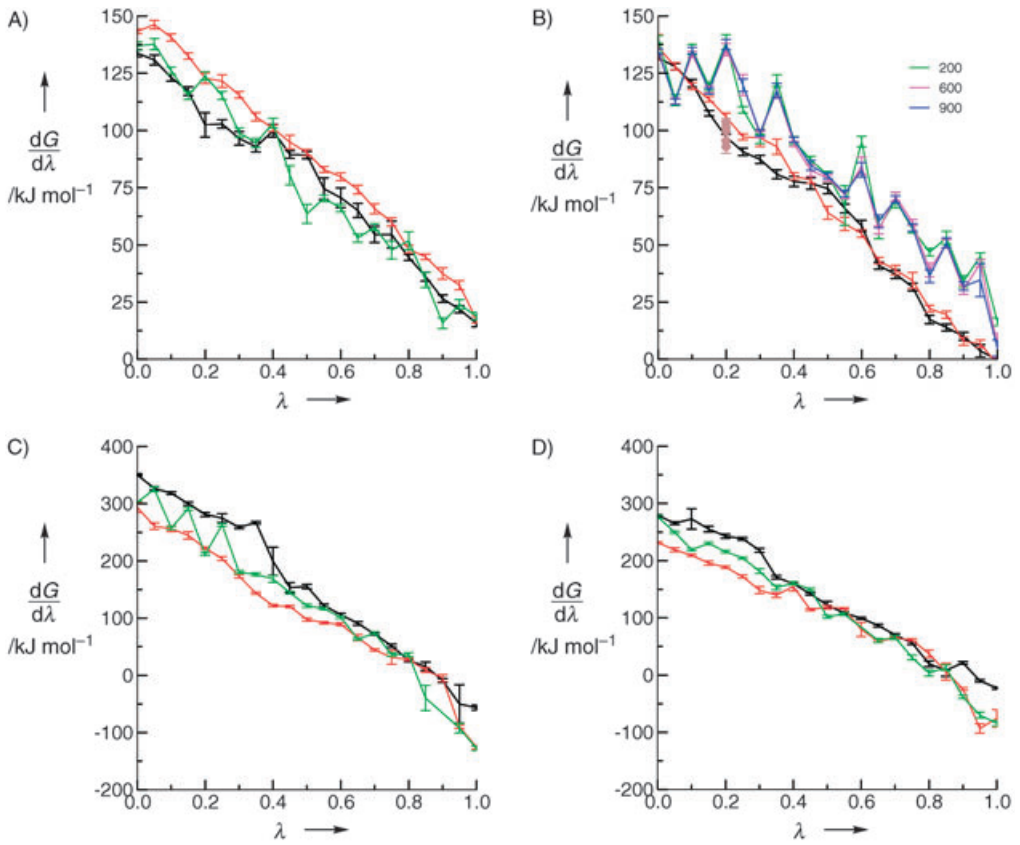

Figure 2. *dG/dλ from Equation (2) as a function of λ for the four legs of the thermodynamic cycle (4). Change between the reduced state (λ = 0) and the oxidised state (λ = 1) at A) pH 5 and B) pH 9. Change between the protonated state of His35 (pH 5, λ = 0) and the deprotonated (neutral) state of His35 (pH 9, λ = 1) for C) the reduced and D) the oxidised protein. The values in black represent a (forward) sequence of 21 equidistant λ values from λ = 0 to 1. The values in red were obtained by extending the simulations at λ = 1 and then using a (backward) sequence of 21 equidistant λ values from λ = 1 to λ = 0. The simulations in the end points, λ = 0 and λ = 1, were equilibrated for 450 ps and sampled for 400 ps; at the intermediate λ values, the equilibration was 50 ps and the sampling was 400 ps. The values in green were obtained through a randomly chosen sequence of λ values. In B, results for different sampling times (red: 200, violet: 600, green: 900 ps) for the random λ sequence are displayed too. The purple dots represent values obtained by using different initial atomic velocities (equilibration 10 ps, sampling 90 ps). The error bars are obtained by block averaging over 400 ps periods (sequential λ changes) and 300 ps (random λ changes).*

hysteresis is the largest, the side chain of His35 becomes deprotonated, and its charge distribution and hydrogen-bond-forming ability change drastically (Table 2). The value of $\langle \partial H/\partial\lambda \rangle_\lambda$ appears correlated to the presence or absence of a hydrogen bond between $H_{\delta 1}$ of His35 and the carbonyl oxygen of Pro36. This hydrogen bond should be present at pH 5, but at pH 9 is replaced by a hydrogen bond between the NH of Gly37 and the N_δ of His35 for large λ values. For this to occur, the peptide plane of Pro36/Gly37 must flip by 180°, a conformational change that is observed in X-ray diffraction studies[3] and that was also observed in the simulations. However, NMR studies suggest that the structural relaxation around His35 is slow[30] and, thus, it is not surprising that insufficient sampling at intermediate λ values is observed.

Experimentally,[8–10] the value for $\Delta G_{ox,red}$(pH 5)$-\Delta G_{ox,red}$(pH 9) is 7 kJ mol^{-1}. The thermodynamic integration calculations yield 15 kJ mol^{-1} (Figure 3A and B) when using sequential λ values and -7 kJ mol^{-1} (Figure 3C and D) when using random λ values along the pathway from the reduced to the oxidised state. When using the two other (vertical) legs of the thermodynamic cycle to obtain this ΔG value, we find 5 (Figure 3A) and 8 kJ mol^{-1} (Figure 3B) using sequential λ values, and 10 (Figure 3C) and 12 kJ mol^{-1} (Figure 3D) using random λ values. On average, these eight values are only 1 kJ mol^{-1} lower than the experimental one, but their spread (22 kJ mol^{-1}) is much larger than 1 kJ mol^{-1}.

The sensitivity of the ΔG values obtained to the particular charge distributions used is illustrated in Figure 3. The results of panels A–D were obtained with the charge set I of Table 2 and those of panel E with the slightly different charge set II. The changes, which only concern the reduced state, yield slightly lower values for $\Delta G_{ox,red}$(pH 5)$-\Delta G_{ox,red}$(pH 9): 4 kJ mol^{-1}, and 8 kJ mol^{-1} obtained via the vertical legs of the cycle. These are comparable to the experimental value of 7 kJ mol^{-1}.

Summarising, we may conclude that the errors in the different ΔG values range from 1 to 10 k_BT, depending on the type of process and choice of sequence of λ values. The cycle-closure value is smaller when using sequential λ values than when choosing a random sequence. In the former case, compensation of errors for the forward and backward processes is likely. A small change of the Hamiltonian causes changes in the relative ΔG values of 1 to 6 k_BT. Extending the sampling times beyond a few hundred picoseconds, doubling the

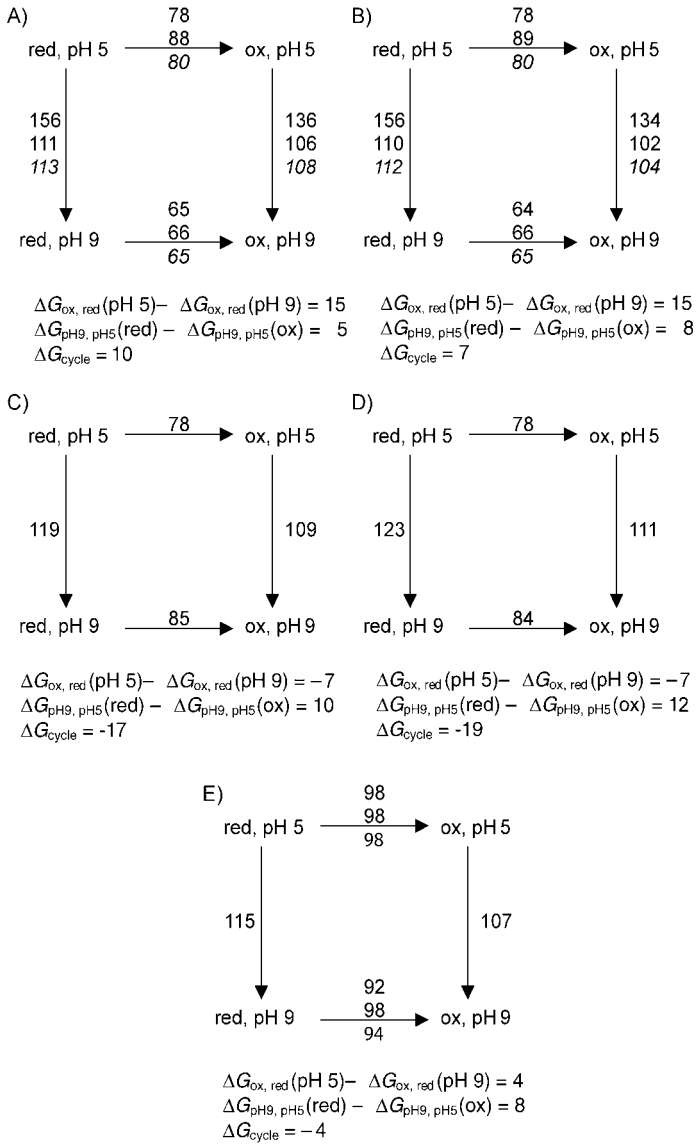

Figure 3. *Free energy differences ΔG [kJ mol^{-1}] along the four legs of the thermodynamic cycle (4) calculated from Equation (3) and obtained by using different sequences of λ values, different equilibration and sampling periods and slightly different Hamiltonians. Sequential λ values: A, B and E. Random λ values: C and D. Δλ = 0.1: A, C and E. Δλ = 0.05: B and D. Standard Hamiltonian (charge set I in Table 2): A, B, C and D. Alternative Hamiltonian (charge set II in Table 2): E. For further explanation, see Figure 2 legend.*

Table 2. *Atomic-charge distributions at the copper site in the reduced and oxidised states and at His35 at pH 5 and pH 9. Charge sets I and II were obtained from DFT calculations with slightly different input structures, as found in PDB[5] files 4azu and 5azu.*

residue	atom	charge (*e*) reduced (I)	reduced (II)	oxidised (I/II)
	Cu	0.215	0.215	0.333
Gly45	C_α	0.124	0.140	0.147
	C	0.381	0.362	0.400
	O	−0.506	−0.518	−0.526
His46	N	−0.730	−0.721	−0.705
	H	0.368	0.380	0.384
	C_α	0.461	0.451	0.429
	C_β	0.094	0.088	0.117
	C_γ	0.077	0.078	0.079
	$N_{\delta1}$	−0.373	−0.384	−0.386
	$C_{\delta2}$	0.014	0.058	0.055
	$C_{\varepsilon1}$	0.350	0.297	0.378
	$N_{\varepsilon2}$	−0.565	−0.426	−0.534
	$H_{\varepsilon2}$	0.479	0.443	0.494
Cys112	C_α	−0.039	−0.035	−0.013
	C_β	0.056	0.071	0.153
	S_γ	−0.629	−0.646	−0.222
His117	C_α	−0.035	−0.039	−0.017
	C_β	0.214	0.180	0.217
	C_γ	−0.003	0.011	0.001
	$N_{\delta1}$	−0.378	−0.368	−0.391
	$C_{\delta2}$	0.142	0.106	0.181
	$C_{\varepsilon1}$	0.352	0.338	0.366
	$N_{\varepsilon2}$	−0.489	−0.441	−0.451
	$H_{\varepsilon2}$	0.409	0.388	0.433
Met121	C_α	−0.110	−0.122	−0.087
	C_β	0.168	0.166	0.156
	C_γ	0.084	0.070	0.131
	S_δ	−0.259	−0.241	−0.261
	C_ε	0.129	0.108	0.141
		charge (*e*) pH 5	pH 9	
His35	C_γ	−0.05	0.13	
	$N_{\delta1}$	0.38	−0.58	
	$H_{\delta1}$	0.30	–	
	$C_{\delta2}$	0.0	0.0	
	$C_{\varepsilon1}$	−0.24	0.26	
	$N_{\varepsilon2}$	0.31	0.0	
	$H_{\varepsilon2}$	0.30	0.19	

number of λ values beyond about 10 or varying the initial velocities yields variations in the ΔG values of only $1\,k_BT$. The slow structural relaxation of the protein in response to a change of λ appears to be the major determinant of the precision of the calculated ΔG value upon a change in pH or redox state.

Since the first applications of the thermodynamic integration[26] and perturbation[31] formulae in MD simulation studies to obtain relative free-energy differences,[32,33] a variety of technical aspects and pitfalls have been identified[34–36] and are still being investigated.[28,37–41] As we find in this study, the most important factors determining the precision of a ΔG value are the choice of pathway or λ-dependence of the Hamiltonian[27,42] and the relaxation of the molecular environment upon changing the Hamiltonian.

Relative redox potentials of azurin mutants

In the simulations, the hysteresis of the forward and reverse redox-state change in azurin is of the order of a few k_BT. Even at this rather low precision, one can still use the thermodynamic-integration approach to predict the relative redox potentials of different mutants of azurin. In Figure 4, the relative free energies $\Delta G_{ox,red}$(pH, mutant)$-\Delta G_{ox,red}$(pH 5, wild-type) calculated from MD simulations are compared to the experimental values.[8–10] The average hysteresis is 6 kJ mol^{-1}. Although some correlation between calculated and measured relative redox potentials is observed, overall the correlation is rather poor. This leads to the conclusion that thermodynamic-integra-

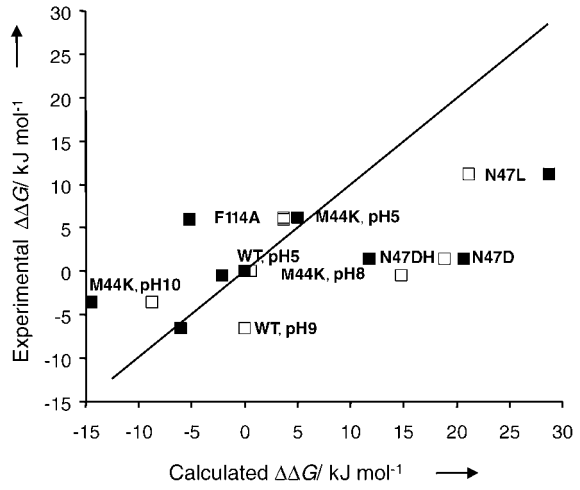

Figure 4. *Comparison of calculated and measured relative redox potentials [kJ mol^{-1}] for azurin and some mutants at different pH values: $\Delta\Delta G = \Delta G_{ox,red}$(pH, mutant)$-\Delta G_{ox,red}$(pH 5, wild type). The experimental values were taken from refs. [42–44] The experimental $\Delta\Delta G$ value for N47L was obtained from the experimental data on* Alcaligenes denitrificans *azurin.[45] The calculated values were obtained by using the forward change ($\lambda = 0$ to 1) from the reduced to the oxidised state (■) and the backward change ($\lambda = 1$ to 0) from the oxidised to the reduced state (□). For the N47DH variant, only the forward calculation was performed ($\Delta\Delta G = 12$ kJ mol^{-1}).*

tion calculations to obtain free-energy differences between different redox states or pH values may only be used to obtain rough estimates of the effects of mutating the protein. Only if ways are found to properly relax and sample the various protein conformations, may the estimates become more than qualitative.

Computational Methods

The MD simulations were carried out by using the GROMACS simulation software[47,48] and the GROMOS96 biomolecular force field, parameter set 43A2.[49–51] Aliphatic CH_n groups were treated as united atoms. The simple-point-charge (SPC) model[52] was used for water.

The atomic charges for the copper and its ligands (Table 2) were obtained from density-functional (DFT) quantum-chemical calculations by using the Amsterdam Density Functional (ADF) software.[53] The positions of the non-hydrogen atoms in the Cu site were taken from the X-ray structure of oxidised azurin at pH 5 (PDB[5] code 4AZU). The Becke–Perdew exchange-correlation functional was used with a triple-zeta valence plus polarisation basis set of Slater-type orbitals. Multipole-derived charge analysis was used to obtain the atomic charges.[54] To obtain charges for the aliphatic united atoms, the charges of the hydrogens were added to those of the carbons. The charges of charge set II differ slightly from those of charge set I and were used to check the effect of changes in the charge distribution on the outcome of the calculations.

The initial coordinates of the atoms in azurin were taken from the X-ray structure of the oxidised form at pH 5. Polar and aromatic hydrogen atoms were added according to standard geometries. His83 was (singly) protonated at $N_{\varepsilon 2}$. The different protonation states of the copper ligands and of His35 in the four simulations (oxidised versus reduced, pH 5 versus pH 9) are shown in Table 2. Since a classical treatment of the interactions of the copper ion with its ligands is a rather poor approximation of the quantum-chemical nature of the latter and may lead to unrealistic deformations of the copper–ligand geometry, the Cu atom was connected to the protein by constraining the distances from the Cu atom to the five coordinating ligands to the respective values in the X-ray structure of the wild-type protein (see Table 3, set A). Although the details of the force field might affect the absolute value of the redox potential, we expect the potential *differences* to be insensitive to them. Because of these constraints, no nonbonded interaction between the copper and its five coordinating atoms was taken into account. The distance constraint set B was taken from the XRD structure of the F114A azurin variant[46] and was used in the MD simulations of this variant. In the MD simulations of the reduced state, one of the $H_{\delta 2}$ atoms of the side chain of Asn47 tends to form a hydrogen bond with the C_γ atom of Cys112, thereby disrupting the coordination sphere of the copper. To prevent this, a repulsive distance restraint was applied between the $N_{?\delta 2}$ atom of Asn47 and the C_γ atom of Cys112 (Table 3).

Table 3. *Bonding parameters for the copper–ligand distance restraining. Set B was only used for the mutant F114A.*

Atom pair	Distance restraint [nm]	
	set A	set B
Cu–O Gly45	0.2955	0.3066
Cu–$N_{\delta 1}$ His46	0.2064	0.2208
Cu–S_γ Cys112	0.2267	0.2215
Cu–$N_{\delta 1}$ His117	0.1978	0.2362
Cu–S_δ Met121	0.3164	0.3006
$N_{\delta 2}$ Asn47–S_γ Cys112	0.45	0.45

The protein was placed in a rectangular periodic box with edge lengths such that the distance between any protein atom and the walls of the box was larger than 0.6 nm. This resulted in a box volume of 27.7 m^3. The remaining space was then filled with 3208 water molecules taken from an equilibrium distribution of liquid water at ambient temperature and pressure. The water configuration was relaxed for 50 steps of energy minimisation, while keeping the protein atoms fixed. Next protein and water degrees of freedom were relaxed through another 50 steps of energy minimisation.

The initial atomic velocities were taken from a Maxwellian distribution at 300 K. The simulations were performed at constant temperature (300 K) and pressure (1 atm) by separately coupling the protein and the solvent to external temperature and pressure baths[55] with coupling times of 0.1 ps for the temperature and 1 ps for the pressure. The pressure was calculated with a molecular virial and held constant with an isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$. In the azurin simulations of the thermodynamic cycle in which charge set I was used, the bond lengths in the protein were kept rigid by using the LINCS algorithm[56] with a geometric tolerance of 10^{-4}. The geometry of the water molecules was kept fixed in the same manner. The equations of motion were integrated by using the leap-frog algorithm and a time step of 2 fs. In the simulations of azurin with charge set II and of the azurin mutants, the bonds in the protein were not constrained and a time step of 1 fs was used.

The nonbonded interaction between atoms in so-called charge groups[49] was calculated according to a spherical triple-range cut-off scheme. Short-range van der Waals and electrostatic interactions were evaluated at every time step by using a charge-group pair list that was generated with a short-range cut-off radius of 0.8 nm between centres of geometry of the protein charge groups and the oxygen atoms of the water molecules. Longer-range van der Waals and electrostatic interactions, between pairs at a distance longer than 0.8 nm and shorter than a long-range cut-off of 1.4 nm, were evaluated every fifth time step, at which point the pair list was also updated, and were kept unchanged between these updates. To approximate the electrostatic interactions beyond the long-range cut-off, a Poisson–Boltzmann reaction-field force was used. The value for the dielectric permittivity of the continuum outside the long-range cut-off was 78.

The MD simulations of the thermodynamic cycle (4) were initialised in the reduced form at pH 5. First, the upper and left legs were simulated forward to the oxidised form (at pH 5) and to pH 9 (in the reduced form), respectively, and then back to the reduced form at pH 5. Second, the lower and right legs were simulated, starting at the reduced form at pH 9 and at the oxidised form at pH 5, respectively, forward to the oxidised form at pH 9, and then back to the respective initial states. At each of the four end points of the thermodynamic cycle, the system was equilibrated for 450 ps and sampled for another 400 ps. The change in λ along the horizontal and vertical pathways was carried out in 10 or 20 steps of $\Delta\lambda = 0.1$ or $\Delta\lambda = 0.05$. At each λ value, 50 ps equilibration was followed by 400 ps sampling. For some cases, the change in λ values was non-sequential. The following random sequence of 21 λ values was used: 0.35, 0.6, 0.2, 0.7, 0.8, 0.55, 0.45, 1.0, 0.05, 0.25, 0.4, 0.85, 0.75, 0.1, 0.95, 0.65, 0.0, 0.15, 0.3, 0.5, 0.9. For each λ value equilibration covered 700 ps and sampling the next 300 ps. The results of these simulations are presented in Table 1 and Figures 2 and 3.

The simulations of the mutants were based on charge set II (Table 2) and involved different numbers of water molecules: 2796 for F114A, 3189 for M44K at pH 5, pH 8 and pH 10, and 4152 for wild-type azurin at pH 5 and pH 9, for N47L, for N47D and for N47DH (the protonated form of Asp47). In these simulations, eight to 13 λ values are used per change of oxidation state. The equilibration at the end points of the legs of the cycle was reduced to 250 ps, and the sampling time per λ value was reduced to 200 ps. For the simulations of the M44K variant, the state of protonation of H35 and K44 was chosen as follows: pH5: both residues protonated; pH 8: H35 deprotonated, K44 protonated; pH 10: both residues deprotonated.

Atomic positions were saved every 0.5 ps for analysis. Hydrogen bonds were calculated by using a geometric criterion. A hydrogen bond is defined by a minimum donor-hydrogen-acceptor angle of 135° and a maximum hydrogen–acceptor distance of 0.25 nm.

Keywords: free-energy calculation · GROMOS force field · molecular dynamics · simulations · thermodynamics

[1] U. Kolczak, C. Dennison, A. Messerschmidt, G. W. Canters in *Handbook of Metalloproteins* (Eds.: A. Messerschmidt, R. Huber, T. Poulos K. Wieghardt), Wiley, Chichester, **2001**, pp. 1170 – 1194.
[2] E. T. Adman, L. H. Jensen, *Isr. J. Chem.* **1981**, *21*,8 – 12.
[3] H. Nar, A. Messerschmidt, R. Huber, M. van der Kamp, G. W. Canters, *J. Mol. Biol.* **1991**, *221*, 765 – 772.
[4] M. van der Kamp, G. W. Canters, S. S. Wijmenga, A. Lommer, C. W. Hilbers, H. Nar, A. Messerschmidt, R. Huber, *Biochemistry* **1992**, *31*, 10194 – 10207.
[5] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.
[6] W. B. Shepard, B. F. Anderson, D. A. Lewandoski, G. E. Norris, E. M. Baker, *J. Am. Chem. Soc.* **1990**, *112*, 7817 – 7819.
[7] C. M. Groeneveld, M. C. Feiters, S. S. Hasnain, J. van Rijn, J. Reedijk, G. W. Canters, *Biochim. Biophys. Acta* **1986**, *873*, 214 – 227.
[8] C. S. StClair, W. R. Ellis, H. B. Gray, *Inorg. Chim. Acta* **1992**, *191*, 149 – 155.
[9] G. Battistuzzi, M. Borsari, L. Loschi, R. Righi, M. Sola, *J. Am. Chem. Soc.* **1999**, *121*, 501 – 506.
[10] M. van der Kamp, G. W. Canters, C. R. Andrew, J. Sanders-Loehr, C. J. Bender, J. Peisach, *Eur. J. Biochem.* **1993**, *218*, 229 – 238.
[11] E. I. Solomon, K. W. Penfield, A. A. Gewith, M. D. Lowery, S. E. Shadle, J. A. Guckert, L. B. Lacroix, *Inorg. Chim. Acta* **1996**, *243*, 67 – 78.
[12] B. W. Beck, J. B. Korner, T. Ichiye, *J. Phys. Chem. B* **1999**, *103*, 8006 – 801.
[13] H. Li, S. P. Webb, J. Ivanic, J. H. Jensen, *J. Am. Chem. Soc.* **2004**, *126*, 8010 – 8019.
[14] A. E. Mark, W. F. van Gunsteren, *J. Mol. Biol.* **1994**, *240*, 167 – 176.
[15] C. Arcangeli, A. R. Bizzarri, S. Cannistraro, *Biophys. Chem.* **2001**, *90*, 45 – 56.
[16] M. Pappalardo, D. Milardi, D. M. Grasso, C. La Rosa, *J. Comput. Chem.* **2003**, *24*, 779 – 785.
[17] B. Rizzuti, M. Swart, L. Sportelli, R. Guzzi, *J. Mol. Model.* **2004**, *10*, 25 – 31.
[18] A. R. Bizzarri, *J. Phys. Condens. Matter* **2004**, *16*, R83 – R110.
[19] D. M. Popović, S. D. Zarić, B. Rabenstein, E. Knapp, *J. Am. Chem. Soc.* **2001**, *123*, 6040 – 6053.
[20] D. Bashford, M. Karplus, G. W. Canters, *J. Mol. Biol.* **1988**, *203*, 507 – 510.
[21] M. S. Formaneck, G. Li, X. Zhang, Q. Cui, *J. Theor. Comput. Chem.* **2002**, *1*, 53 – 67.
[22] G. Li, X. Zhang, Q. Cui, *J. Phys. Chem. B.* **2003**, *107*, 8643 – 8653.
[23] U. Ryde, *Curr. Opin. Chem. Biol.* **2003**, *7*, 136 – 142.
[24] S. Moon, S. Patchkowskii, D. R. Salahub, *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 287 – 295.
[25] M. H. M. Olsson, G. Hong, A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 5025 – 5039.
[26] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300 – 313.
[27] W. F. van Gunsteren, X. Daura, A. E. Mark, *Helv. Chim. Acta* **2002**, *85*, 3113 – 3129.
[28] D. A. Hendrix, C. Jarzynski, *J. Chem. Phys.* **2001**, *114*, 5974 – 5981.
[29] W. Yang, R. Bitetti-Putzer, M. Karplus, *J. Chem. Phys.* **2004**, *120*, 2618 – 2628.
[30] A. P. Kalverda, M. Ubbink, G. Gilardi, S. S. Wijmenga, A. Crawford, L. J. C. Jeuken, G. W. Canters, *Biochemistry* **1999**, *38*, 12690 – 12699.
[31] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420 – 1426.
[32] J. P. M. Postma, H. J. C. Berendsen, J. R. Haak, *Faraday Symp. Chem. Soc.* **1982**, *17*, 55 – 67.
[33] M. Mezei, D. L. Beveridge, *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1 – 23.
[34] W. F. van Gunsteren in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, (Eds.: W. F. van Gunsteren, P. K. Weiner), Escom Scientific, Leiden, **1989**, pp. 27 – 59.
[35] W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, P. E. Smith in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, *Vol.* 2 (Eds.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson) Escom, Leiden, **1993**, pp. 315 – 348.
[36] Y. Y. Shi, A. E. Mark, C. X. Wang, F. Huang, H. J. C. Berendsen, W. F. van - Gunsteren, *Protein Eng.* **1993**, *6*, 289 – 295.
[37] N. Lu, D. A. Kofke, *J. Chem. Phys.* **2001**, *114*, 7303 – 7311.
[38] N. Lu, D. A. Kofke, *J. Chem. Phys.* **2001**, *115*, 6866 – 6875.
[39] N. Lu, D. A. Kofke, T. B. Woolf, *J. Comput. Chem.* **2004**, *25*, 28 – 40.
[40] N. Lu, D. A. Kofke, T. B. Woolf, *J. Phys. Chem. B.* **2003**, *107*, 5598 – 5611.
[41] S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, *J. Chem. Phys.* **2003**, *119*, 3559 – 3566.
[42] A. E. Mark, W. F. van Gunsteren, H. J. C. Berendsen, *J. Chem. Phys.* **1991**, *94*, 3808 – 3816.
[43] L. Sjölin, L. Tsai, V. Langer, T. Pascher, G. Karlsson, M. Nordling, H. Nar, *Acta Crystallogr. D* **1993**, *49*, 449 – 457.
[44] M. van de Kamp, G. W. Canters, C. R. Andrew, J. Sanders-Loehr, C. J. Bender, J. Peisach, *Eur. J. Biochem.* **1993**, *218*, 229 – 238.
[45] L. Tsai, L. Sjölin, V. Langer, T. Pascher, H. Nar, *Acta Crystallogr. D* **1995**, *51*, 168 – 176.
[46] C. W. G. Hoitink, G. W. Canters, *J. Biol. Chem.* **1992**, *267*, 13836 – 13842.

[47] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Communications* **1995**, *91*, 43 – 56.
[48] E. Lindahl, B. Hess, D. van der Spoel, *J. Mol. Model.* **2001**, *7*, 306 – 317.
[49] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag, Zürich, Switzerland **1996**, pp. 1 – 1042.
[50] X. Daura, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **1998**, *19*, 535 – 547.
[51] L. D. Schuler, W. F. van Gunsteren, *Mol. Simul.* **2000**, *25*, 301 – 319.
[52] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in *Intermolecular Forces* (Ed.: B. Pullman), Reidel, Dordrecht, **1981**, pp. 331 – 342.
[53] G. T. Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. Van Gisbergen, J. G. Snijders, T. Ziegler, J . *Comput. Chem.* **2001**, *22*, 931 – 967.
[54] M. Swart, P. T. van Duijnen, J. G. Snijders, *J. Comput. Chem.* **2001**, *22*, 79 – 88.
[55] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684 – 3690.
[56] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463 – 1472.

Received: July 15, 2004
Published online on March 3, 2005

