# A Computational Study of the Open and Closed Forms of the N-Lobe Human Serum Transferrin Apoprotein

David Rinaldo and Martin J. Field
Laboratoire de Dynamique Moléculaire, Institut de Biologie Structurale Jean-Pierre Ebel, Commissariat à l'Energie Atomique, and the Centre National de Recherche Scientifique, Grenoble, France

ABSTRACT Human serum transferrin tightly binds ferric ions in the blood stream but is able to release them in cells by a process involving receptor-mediated endocytosis and decrease in pH. Iron binding and release are accompanied by a large conformation change. In this study, we investigate theoretically the open and closed forms of the N-lobe human serum transferrin apoprotein by performing $pK_a$ calculations and molecular dynamics and free-energy simulations. In agreement with the hypothesis based on the x-ray crystal structures, our calculations show that there is a shift in the $pK_a$ values of the lysines forming the *dilysine trigger* when the conformation changes. We argue, however, that simple electrostatic repulsion between the lysines is not sufficient to trigger domain opening and, instead, propose an alternative explanation for the dilysine-trigger effect. Analysis of the molecular dynamics and free-energy results indicate that the open form is more mobile than the closed form and is much more stable at pH 5.3, in large part due to entropic effects. Despite a lower free energy, the dynamics simulation of the open form shows that it is flexible enough to sample conformations that are consistent with iron binding.

## INTRODUCTION

The transferrins are a family of homologous iron-binding glycoproteins that are present in all vertebrates and some insects. Their main function is to control the levels of free iron in body fluids by binding and sequestering $Fe^{3+}$. They also prevent the formation of ferric hydroxides and protect against the toxic effects of iron which can catalyze the formation of free radicals that cause cell damage (Sun et al., 1999). In the transferrin family, lactoferrin (Lf) is widespread in a variety of secretory fluids such as milk, tears, bile, pancreatic juice, mucosal fluid, and white blood cells, whereas ovotransferrin (oTf) is mainly present in egg white. They both act as bacteriostatic agents because they are able to bind iron so tightly that it is unavailable for bacterial growth (Brock, 2002; Baker, 1994; Feeney and Komatsu, 1966). Serum transferrin (sTf) has the role of transporting iron from sites of absorption to cells. At the cell surface, the iron-loaded sTf is bound by the transferrin receptor (TfR) and is internalized via receptor-mediated endocytosis. The iron is released as the pH of the vesicle decreases from 7.4 to 5.5 with sTf remaining bound to the receptor. Afterwards, iron is transported out of the vesicle and is stored by ferritin whereas sTf and its receptor are recycled. Each sTf molecule is able to undergo ~100 such cycles before it is degraded (Zak et al., 2002).

Human sTf consists of a single polypeptide chain of 679 amino acids with a molecular weight of ~80 kDa. It is folded into two globular lobes, the N-lobe ($sTf_N$) and the C-lobe ($sTf_C$), with a short peptide chain connecting them. The two lobes are both able to bind one ferric iron and do not interact strongly even though there is some evidence for interactions between them (Brock, 1985). Each lobe can further be subdivided into two domains, which form a deep cleft where the iron binding site is housed. Furthermore, there are strong homologies in the amino acid sequences not only between transferrins (60% between sTf and Lf) but also between the two lobes (40%). This has led some authors to conclude that the protein evolved via gene duplication (MacGillivray and Brew, 1975; Gorinsky et al., 1979; Williams, 1982).

The iron binding sites are similar in both lobes and among family members. The iron coordination is distorted octahedral with ligands provided by four amino acid side chains—two tyrosines, one histidine, and one aspartate. The coordination is completed by an anion which, in vivo, is usually (bi)-carbonate. One striking aspect of the chemistry of transferrins is that binding of iron needs concomitant binding of this anion which is, therefore, called the synergistic anion (Baker, 1994).

Although the two lobes are homologous, they have subtle differences in their properties. $sTf_N$ begins to release iron at pH ~ 5.7, whereas the $sTf_C$ retains iron until pH ~ 4.8. There are also differences between transferrin family members as, for example, Lf retains iron to as low as pH 3.5 (Baker et al., 2002). It is thought that $sTf_N$ has a specific molecular mechanism for iron release at pH 5.7. Two lysines (Lys-206 and Lys-296) have been proposed to play this role as they are on opposite faces of the iron-binding cleft and are thought to be bridged by a single proton in the closed form of the protein. At medium and high pH, these two lysines would help to hold the two domains together, whereas at low pH (<6) the lysines would both be protonated, causing the two domains to repel each other. Domain opening would be followed by iron release as the iron becomes exposed to solvent (Dewan et al., 1993) although, as noted in MacGillivray et al. (1998), this step is probably preceded by other protonation events.

*Submitted January 28, 2003, and accepted for publication July 24, 2003.*

Address reprint requests to Martin J. Field, Laboratoire de Dynamique Moléculaire, Institut de Biologie Structurale Jean-Pierre Ebel, 41 rue Jules Horowitz, 38027 Grenoble cedex 1, France. Tel.: 33-43-878-5919; Fax: 33-43-878-5122; E-mail: mjfield@ibs.fr.



0006-3495/03/12/3485/17 $2.00

Although the principal biochemical function of transferrins is to reversibly complex iron, they can also bind a very wide variety of other essential and toxic metals (>30) (Taylor, 1993). Transferrins could therefore play an important role in metal toxicity processes and be implicated in several diseases such as Alzheimer's or leukemia (Yajima et al., 2000). Hence, it is important to understand why they are able to bind metals so tightly and how they are able to release them.

To help gain further insight into these processes, we are studying $sTf_N$ with a range of computational tools. In this article, we present results of calculations that we have performed for the apoprotein in both its open and closed conformations. We have carried out $pK_a$ calculations to investigate the protonation state of residues in the protein, molecular dynamics simulations to investigate the dynamics of the two forms, and free-energy calculations to determine their relative stabilities. The methods that we used and the results that we obtained are outlined in the following sections, although their implication for the function of sTf is left until the end of the article.

## METHODS

### $pK_a$ calculations

The protocol that we used for the $pK_a$ calculations follows that described by Gilson and co-workers (Gilson, 1993; Antosiewicz et al., 1994) and by David (1996). To briefly summarize, the method determines the $pK_a$ values for relevant ionizable groups in the system by calculating how the environment perturbs the intrinsic $pK_a$ of these groups. The environmental effect is assumed to be entirely electrostatic and is determined by solving the linearized Poisson-Boltzmann equation for the electrostatic potential at each group and for various protonation states of the protein. Although the method that we use is relatively standard, we have made some changes, particularly in how the protonation of histidine residues is evaluated. We intend to describe these developments in more detail in a future publication.

For the $pK_a$ calculations, we used the highest-resolution crystallographic structures of the apo- and holoforms of $sTf_N$. The structures had Protein Data Bank (see http://www.rcsb.org/pdb/cgi/queryForm.cgi) codes of 1bp5 (Jeffrey et al., 1998) at 2.20 Å resolution and 1a8e (MacGillivray et al., 1998) at 1.60 Å resolution, respectively. All water molecules were removed from both structures. For the apoform, the D-chain in the crystal cell was chosen but residues 1–3 were missing, whereas residues 1–2 and 332–337 were not available in the holoform structure. To be able to compare results, it was decided to map the holoform onto the apoform. Residue 3 of the holoform was therefore removed and residues 332–337 of the apoform were added to the holoform by superimposing the carbon and oxygen backbone atoms of residue 331 of the holo- and apoforms. These structures served both for our $pK_a$ calculations and as the starting point for our molecular dynamics and free-energy simulations (see below).

We have automated the calculation of $pK_a$ values for any protein, including the setup part, using the VMD protein visualization program as the graphical interface (Humphrey et al., 1996). The procedure involves adding the coordinates of polar hydrogens to a protein coordinate file using the molecular modeling program CHARMM (Brooks et al., 1983) followed by calculation of the electrostatic potentials using the UHBD program and a four-step method of electrostatic focusing for increased accuracy (Briggs et al., 1989). The multisite $pK_a$ calculation method of Gilson (1993) was used. All electrostatic potential calculations were done at 300 K, with an ionic strength of 150 mM. The solvent-accessible region was defined as the volume that could be swept out by a spherical probe of radius 1.4 Å whereas the radius of ions was set to 2.0 Å. Dielectric constants of 80 and of 20 were used for the solvent and the protein, respectively. This last value, although large, is consistent with our experience with the method and with previous protein $pK_a$ calculations (Antosiewicz et al., 1994) and was found to give reasonable results in a study of Lf $pK_a$ values (Lee and Goodfellow, 1998). Physically, it is a way of mimicking the internal flexibility and polarizability of the protein which otherwise is not taken into account. The grid sizes for the electrostatic focusing were 5.0, 1.0, 0.5, and 0.25 Å, respectively. All atomic charges and radii used were taken from the UHBD program except for the iron charge (+3) and the charges on the carbonate, the values for which were taken from the quantum chemical calculations of Lee and Goodfellow (1998).

### Molecular dynamics simulations

Two simulations were performed on $sTf_N$—one on the open apoform, and the other on the closed form but without iron and carbonate. For starting structures we used the same ones as for the $pK_a$ calculations, except that iron and carbonate were removed from the closed form. The protonation states of the residues in the proteins were chosen so as to follow those calculated for the open form of the apoprotein at a pH of 5.3 (which is the crystallization pH of the apoprotein structure; Jeffrey et al., 1998). This meant that most acidic and basic residues had their normal protonation state whereas His-14, His-25, His-242, His-249, His-273, and His-289 were protonated, His-119 was neutral with the proton on ND1, and His-207 and His-300 were neutral with protons on NE2. The only differences between the protonation states of the open and closed forms of the protein were for residues Glu-237, His-249, and His-289. Glu-237, which has a $pK_a$ of 5.3 in the open form, was considered as protonated, but this residue does not seem to be critical in our study as it is located on the surface of the protein. The $pK_a$ values for His-289 straddle the pH value of 5.3, but otherwise differ little, so mixtures of protonated and unprotonated residues could be expected in each form. This left His-249 which, because it occurs in the iron-binding site, undergoes a large $pK_a$ change on domain closure.

After assignment of the protonation states, all hydrogens—both polar and nonpolar—were added to the protein using CHARMM (Brooks et al., 1983) and their positions were minimized. This was followed by a minimization of the side chains, with the backbone fixed, and then of the whole protein. The minimizations were carried out with harmonic constraints on the moveable atoms, the force constants for which were initially high but were gradually reduced to zero. Once minimization was complete, the system was immersed in an equilibrated water box of volume $80^3$ Å$^3$. All water molecules within 2.8 Å from the protein were removed. Five chloride ions were also randomly added to the water box so as to ensure that the total charge of the system was zero. Each simulation system consisted of ~49,000 atoms. Water molecules within 10 Å of the protein and ions were first minimized and then equilibrated at 300 K with a short 1-ps Langevin dynamics simulation. The timestep for the simulation was 1 fs, the geometries of the water molecules were fixed using SHAKE, a friction coefficient of 50 ps$^{-1}$ was employed for each atom, and the electrostatic and Lennard-Jones interactions were calculated with the particle-mesh Ewald method and a 12 Å cutoff, respectively. The equilibration of the waters was followed by an equilibration of the whole system for 5 ps with the same algorithm.

The simulation proper was performed at constant pressure and temperature using the default CHARMM algorithm with parameters of 1 atm for the pressure, with a piston mass of 500 atomic mass units and a piston collision frequency of 50 ps$^{-1}$, and a temperature of 300 K, with a thermal piston mass of 1000 kcal $\times$ mol$^{-1}$ $\times$ ps$^2$. Initially 20 ps of equilibration was performed but, in the end, it was found that 520 ps were necessary. The equilibration was followed by a simulation of 1.5 ns during which data were collected every 0.5 ps. All calculations were performed with the CHARMM (Brooks et al., 1983) program and its force field for proteins (MacKerell Jr. et al., 1998). Subsequent analyses were performed with CHARMM and with the DYNAMO library (Field et al., 2000; Field, 1999).

## Essential dynamics analysis

As part of our analysis of the dynamics trajectories, we used the essential dynamics method whose aim is to extract the important global motions from the dynamics and describe them with a small number of atomic displacement vectors which form the so-called *essential subspace* (Amadei et al., 1993; van Aalten et al., 1995; Nolde et al., 2002). To perform the analysis, only protein atoms were considered. First, global translations and rotations were removed from the structures forming the trajectories by superimposing the nonhydrogen atoms of each structure on the first structure of the last 1.5 ns of each simulation. Subsequently, a variance-covariance matrix was constructed from the resulting trajectories using all heavy atoms. This matrix is defined by Krzanowski (1988) as

$$\sigma = \mathrm{cov}(\mathbf{x}) = \frac{1}{N-1}\sum_{t=0}^{N}(\mathbf{x_t} - \langle \mathbf{x_t} \rangle_t)(\mathbf{x_t} - \langle \mathbf{x_t} \rangle_t)^T, \quad (1)$$

where $\mathbf{x_t}$ is the structure coordinate vector at time $t$, $\langle\ \rangle_t$ denotes an average over time, and $T$ is the transpose operator. The symmetric matrix $\sigma$ can always be diagonalized by an orthogonal coordinate transformation. The resulting eigenvectors define a new set of coordinates and their eigenvalues are the mean square fluctuations the new coordinates generate. The eigenvectors can be viewed as directions along which variance is maximum (Krzanowski, 1988) whereas the eigenvalues indicate the amplitudes of motion in these directions. Construction of a plot of eigenvector indices against eigenvalues, where eigenvectors are sorted by decreasing eigenvalue, shows that there are only few eigenvectors with large eigenvalues and that motions can be well-depicted by just analyzing a few variables (typically 5–20). The major assumption with essential dynamics is that the relevant motions for the function of the protein are described by eigenvectors with large eigenvalues. All the essential dynamics analysis was performed using the DYNAMO library (Field, 1999; Field et al., 2000).

## MM-PBSA free-energy calculations

To estimate free-energies of the proteins we used the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method that has been mainly developed by Kollman and co-workers (Kollman et al., 2000; Wang et al., 2001). The method involves taking structures from a molecular dynamics simulation, removing solvent and counterions, and then determining the free energy using the equation of Kollman et al. (2000),

$$\overline{G} = \overline{E}_{MM} + \overline{G}_{solv-pol} + \overline{G}_{solv-non-pol} - TS, \quad (2)$$

where $\overline{G}$ is the average free energy, $\overline{E}_{MM}$ is the average molecular mechanical energy of the isolated solute, $\overline{G}_{solv-pol}$ and $\overline{G}_{solv-non-pol}$ are the solvation free energy calculated by solving the Poisson-Boltzmann equation corrected with a simple nonpolar free-energy surface area term, and $S$ is the solute entropy. Snapshots were extracted from the dynamics trajectories every 5 ps and the MM energy of the isolated solute was calculated using both the CHARMM force field and the OPLS-AA force field (Jorgensen et al., 1996) that is implemented in DYNAMO. Solvent-accessible surface areas were estimated for the same snapshots with MSMS (Sanner et al., 1996) using PARSE radii (Sitkoff et al., 1994) and the nonpolar contributions to solvation free energy were calculated using the formula

$$\overline{G}_{solv-non-pol} = \gamma A + b, \quad (3)$$

where $A$ is the surface area, $\gamma$ is taken to be 0.00542 kcal $\times$ mol$^{-1}$ $\times$ Å$^{-2}$, and $b$ is 0.92 kcal $\times$ mol$^{-1}$ (Srinivasan et al., 1998). Polar solvation free energies were evaluated with UHBD (Briggs et al., 1989) using PARSE radii (Sitkoff et al., 1994). A two-step focusing method was applied with a 2 Å spacing and a cubic box with $110^3$ grid points for the first run, and a 0.5 Å spacing and a cubic box with $185^3$ grid points for the second. Following Kollman et al. (2000), the solute was assigned a dielectric constant of 1, and the solvent a dielectric constant of 80. A low dielectric constant is used for the protein because, in contrast to the $pK_a$ calculations, the flexibility of the protein is explicitly taken into account. Equally, it is important to use the same dielectric constant for the determination of the MM internal electrostatic and solvation energies so as to maintain consistency between the two terms. The solvent boundary was taken as the molecular surface defined by a 1.4 Å probe sphere whereas the ionic radius was set to 2.0 Å. CHARMM (MacKerell Jr. et al., 1998), OPLS-AA (Jorgensen et al., 1996), and PARSE (Sitkoff et al., 1994) charges were tested.

The calculation of the entropic term presents some difficulties but is critical, especially in our case, where a conformational change could have an important impact on the entropy of the structure. Instead of the normal mode analyses normally favored by Kollman et al. (2000), we chose to evaluate this term by the use of a quasiharmonic analysis of the dynamics following the work of Schlitter (1993), van Gunsteren and co-workers (Schafer et al., 2000, 2002), and Karplus and co-workers (Karplus and Kushick, 1981; Levy et al., 1984; Tidor and Karplus, 1994; Brooks et al., 1995; Andricioaei and Karplus, 2001). Translational and rotational contributions were neglected. The entropy $S$ of a system is given by the well-known Boltzmann expression

$$S = -k_B \sum_{i=1}^{N} P_i \ln P_i, \quad (4)$$

where $P_i$ is the probability of the system to be in state $i$ (implying $\sum_{i=1}^{N} P_i = 1$) and $k_B$ is the Boltzmann constant. The quasiharmonic approximation assumes that the fluctuations observed in the motions of the system can be described by a normalized multivariate Gaussian probability distribution with a covariance, $\sigma$, given by Eq. 1 as

$$P(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^T \sigma^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle)\right). \quad (5)$$

Diagonalization of the mass-weighted covariance matrix gives eigenvalues, $\lambda_i$, from which the quasiharmonic frequencies $\omega_i = (k_B T/\lambda_i)^{1/2}$ can be obtained (see Andricioaei and Karplus, 2001, for more details). The entropy is derived assuming the same formula as for a quantum harmonic oscillator,

$$S = k_B \sum_{i=1}^{n} \frac{\hbar\omega_i/k_B T}{e^{\hbar\omega_i/k_B T} - 1} - \ln(1 - e^{-\hbar\omega_i/k_B T}), \quad (6)$$

where $n$ is the number of nonzero eigenvalues. In our proteins, there are 2586 nonhydrogen atoms and so, in principle, 7758 ($3 \times 2586$) possible eigenvectors. However, as we could only use 3000 structures for each analysis (those obtained during the last 1.5 ns of our simulations), only 3000 of the eigenvalues were nonzero. For comparison purposes, we also used the formula derived by Schlitter (1993),

$$S < S' = \frac{1}{2} k_B \sum_i \ln\left[1 + \left(\frac{k_B T e^2}{\hbar^2}\right)\lambda_i\right]. \quad (7)$$

# RESULTS

## $pK_a$ calculations

$pK_a$ calculations were performed to determine the protonation state of ionizable residues in various forms of $sTf_N$. The results of the calculations are presented in Table 1 for the open form of the apoprotein, for the closed form of the apoprotein (the holoform with iron and carbonate removed), for the closed apoform with carbonate only, and for the holoform (with both carbonate and iron). As well as the absolute values of the $pK_a$ values for each protonatable residue, the shifts in $pK_a$ values are given using the open

**TABLE 1 The $pK_a$ values for protein residues in various conformations of human $sTf_N$**

| Residue | Apo open | Apo closed | $CO_3^{2-}$ closed | $Fe^{3+}CO_3^{2-}$ closed | $\Delta pK_a^1$ | $\Delta pK_a^2$ | $\Delta pK_a^3$ | $\Delta pK_a^4$ |
|---|---|---|---|---|---|---|---|---|
| Arg-7* | 13.7 | 14.0 | 14.0 | 14.0 | 0.3 | 0.0 | 0.0 | 2.0 |
| **Glu-13** | **3.2** | **2.2** | **2.3** | **2.3** | **−1.0** | **0.1** | **0.0** | **−1.7** |
| **Glu-15*** | **2.0** | **0.0** | **1.2** | **0.0** | **−2.0** | **1.2** | **−1.2** | **−4.0** |
| Lys-18* | 12.8 | 12.8 | 13.0 | 12.8 | 0.0 | 0.2 | −0.2 | 2.4 |
| Asp-24* | 2.3 | 1.5 | 1.5 | 1.5 | −0.8 | 0.0 | 0.0 | −2.5 |
| Tyr-45* | 11.8 | 11.7 | 12.1 | 11.7 | −0.1 | 0.4 | −0.4 | 2.1 |
| Asp-58* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | −4.0 |
| **Asp-63*** | **0.0** | **0.0** | **1.4** | **0.0** | **0.0** | **1.4** | **−1.4** | **−4.0** |
| Asp-69* | 2.2 | 1.6 | 1.8 | 1.8 | −0.6 | 0.2 | 0.0 | −2.2 |
| Glu-83* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | −4.0 |
| Tyr-85* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 4.4 |
| **Tyr-95*** | **13.5** | **14.0** | **14.0** | **1.2** | **0.5** | **0.0** | **−12.0** | **−8.4** |
| **Tyr-96** | **9.5** | **10.9** | **10.9** | **10.9** | **1.4** | **0.0** | **0.0** | **1.3** |
| **His-119*** | **3.3** | **4.4** | **5.1** | **5.0** | **1.1** | **0.7** | **−0.1** | **−1.3** |
| **Arg-124*** | **12.6** | **14.0** | **14.0** | **14.0** | **1.4** | **0.0** | **0.0** | **2.0** |
| **Tyr-136*** | **9.7** | **12.0** | **12.0** | **12.1** | **2.3** | **0.0** | **0.1** | **2.5** |
| Arg-143* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| **Asp-163** | **4.5** | **3.2** | **3.2** | **3.2** | **−1.3** | **0.0** | **0.0** | **−0.8** |
| Tyr-185* | 12.0 | 12.0 | 12.1 | 12.1 | 0.0 | 0.1 | 0.0 | 2.5 |
| **Tyr-188*** | **10.2** | **13.6** | **14.0** | **0.0** | **3.4** | **0.4** | **−14.0** | **−9.6** |
| Asp-197* | 1.5 | 2.3 | 2.3 | 2.3 | 0.8 | 0.0 | 0.0 | −1.7 |
| **Lys-206*** | **11.1** | **6.8** | **8.1** | **7.4** | **−4.3** | **1.3** | **−0.7** | **−3.0** |
| **Hsd-207*** | **1.5** | **0.0** | **0.0** | **0.0** | **−1.5** | **0.0** | **0.0** | **−6.3** |
| **Glu-212*** | **2.8** | **1.1** | **1.6** | **1.4** | **−1.7** | **0.5** | **−0.2** | **−2.6** |
| Arg-232* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| Lys-233* | 12.8 | 12.0 | 12.0 | 12.0 | −0.8 | 0.0 | 0.0 | 1.6 |
| Lys-239* | 12.8 | 12.2 | 12.2 | 12.2 | −0.6 | 0.0 | 0.0 | 1.8 |
| **Asp-240** | **3.8** | **2.8** | **2.9** | **2.8** | **−1.0** | **0.1** | **−0.1** | **−1.2** |
| **His-249** | **6.1** | **0.0** | **8.7** | **0.0** | **−6.1** | **8.7** | **−8.7** | **−6.3** |
| Arg-254* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| Glu-260* | 1.6 | 2.5 | 2.5 | 2.5 | 0.9 | 0.0 | 0.0 | −1.5 |
| Lys-276* | 12.5 | 12.8 | 12.8 | 12.8 | 0.3 | 0.0 | 0.0 | 2.4 |
| **Lys-291** | **11.2** | **12.3** | **12.4** | **12.3** | **1.1** | **0.1** | **−0.1** | **1.9** |
| **Asp-292*** | **1.7** | **0.0** | **0.0** | **0.0** | **−1.7** | **0.0** | **0.0** | **−4.0** |
| **Lys-296*** | **14.0** | **12.0** | **14.0** | **13.6** | **−2.0** | **2.0** | **−0.4** | **3.2** |
| **Hsd-300*** | **4.0** | **2.5** | **2.5** | **2.6** | **−1.5** | **0.0** | **0.1** | **−3.7** |
| **Tyr-314*** | **13.0** | **14.0** | **14.0** | **14.0** | **1.0** | **0.0** | **0.0** | **4.4** |
| **Tyr-317** | **11.4** | **9.5** | **9.5** | **9.5** | **−1.9** | **0.0** | **0.0** | **−0.1** |
| **Tyr-319*** | **10.6** | **11.4** | **13.4** | **12.3** | **0.8** | **2.0** | **−1.1** | **2.7** |
| Arg-324* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| Arg-327* | 14.0 | 14.0 | 14.0 | 14.0 | 0.0 | 0.0 | 0.0 | 2.0 |

Only residues with a $pK_a$ shift between the different forms larger than 1 $pK_a$ unit (*in bold*) or those with a $pK_a$ shift >2 units compared to the reference values (*with asterisk*) are listed. All values are given in the 0–14 pH window. Histidine states were determined at pH 5.3. Hsd is a histidine where the ionizable group is on ND1, i.e., where the proton is on NE2, whereas His is a histidine where the ionizable group is on NE2, i.e., where the proton is on ND1. $\Delta pK_a^1$, $\Delta pK_a^2$, and $\Delta pK_a^3$ are the $pK_a$ shifts for the closed apo, the closed $CO_3^{2-}$ and the closed $Fe^{3+}CO_3^{2-}$ holo forms, respectively, using the open apoform as reference. $\Delta pK_a^4$ is the $pK_a$ shift between the closed $Fe^{3+}CO_3^{2-}$ form and the standard reference $pK_a$ values of the isolated amino acids.

apoform as reference. The final column shows the $pK_a$ shifts between the holoform and the reference values of $pK_a$ for the amino acids. Although the $pK_a$ values for all ionizable groups in the protein were determined, in the interests of brevity, we only give values for residues which display large shifts between different conformations ($\Delta pK_a > 1$) or compared to their standard $pK_a$ values ($\Delta pK_a > 2$). Likewise, in what follows below, we concentrate on a few residues of particular interest and do not discuss all of the $pK_a$ results in the table.

The $pK_a$ values of all residues involved in the coordination of iron (Asp-63, Tyr-95, Tyr-188, and His-249) undergo big $pK_a$ shifts to lower pH when iron is bound. This is not surprising as the presence of $Fe^{III}$ with its charge of $^{+3}$ disfavors the protonation of these ligands. Clearly, Asp-63, Tyr-95, and Tyr-188 will be deprotonated but the case of His-249 needs more discussion. Our calculations show that it is in an unprotonated, neutral state, which is logical as otherwise a proton and the iron would be too close (as in our model the ionizable site is the nitrogen atom coordinated to iron). However, our model will probably be unrealistic, 1), because it takes no account of electronic structure and polarization effects which are likely to be important for a proper description of the iron-histidine coordination; and

2), because our model does not permit formation of negative histidine (for which the $pK_a$ is ~14; El Yazal and Pang, 1999). We treat this point further in the Discussion section.

Asp-63 has a very low $pK_a$ (<0) value in both the open and closed conformations and even in the presence of carbonate. This can be explained by a hydrogen bond between Asp-63 and Lys-296 for the open conformation and between Asp-63 and Ser-125 and Gly-65 for the closed conformation. Tyr-95 has a very high $pK_a$ value (>13.5) in all conformations except when iron is loaded. In the open form, the reason for the high $pK_a$ value is unclear as the residue is exposed to solvent and has only a single, long-range interaction with Tyr-188. In the closed conformation, interactions with Asp-63, Tyr-188 and, when present, carbonate explain the higher $pK_a$ value. Tyr-188 has a normal $pK_a$ value in the open apoform but undergoes a big shift in the closed conformation, in large part due to interactions with Asp-63. The $pK_a$ of Tyr-95 is slightly lower in the closed apoform compared to that of Tyr-188, because of the proximity of the positively charged Lys-296.

Lys-18 has a high $pK_a$ value in all forms because its side chain points to the loops created by residues 275–296 and interacts with several carbonyl oxygens of the protein backbone in this region. This lysine has a striking conformation as it is perpendicular to the axis of helix 1 (the labeling follows the nomenclature employed by Haridas et al., 1995) and is pointed toward the well-conserved structural $\gamma$-turn motif. Using sequence alignment (Altschul et al., 1997), it appears that this lysine and the following residue (Cys-19) are very well-conserved in transferrins, lactoferrins, and ovotransferrins for all organisms for which the sequence is known, except for the C-lobe of a marsupial, the brushtail possum, where a valine seems to be inserted between the lysine and the cysteine. Therefore, Lys-18 could play an important structural role.

Glu-83 and Tyr-85 have low and high $pK_a$ values, respectively, because they are hydrogen-bonded to one another. As noticed by Jeffrey et al., 1998, there is a direct interaction between Glu-83 and His-249 in the open form which is replaced by an interaction with a water molecule in the closed form.

Arg-124 has a higher $pK_a$ value in the closed conformations than the open form. This residue has hydrogen bonds with carbonate, when it is present, but the high $pK_a$ value occurs even in its absence.

Lys-206 exhibits a very low $pK_a$ (<7.5) value in the closed conformation whereas Lys-296, with which it interacts strongly, has a very large $pK_a$ value. These two lysines form the ''dilysine trigger'' which has been hypothesized to exist from examination of the crystallographic structures (Baker and Lindley, 1992; Dewan et al., 1993). Although Lys-296 is closest to the iron atom, it does not have the lower $pK_a$ value, probably because of interactions it has with Tyr-85 and Tyr-188. Experimental support for a higher $pK_a$ value for Lys-296, rather than Lys-206, comes from mutation studies of Arg-210 in Lf (Peterson et al., 2002).

His-207 was found to have the initial proton on the nitrogen atom NE2. This is consistent with the fact that the ND1 atom seems to be hydrogen-bonded with the amide nitrogen of Tyr-96. The $pK_a$ value of His-207 is therefore very low (<2). His-207 also interacts with Lys-206 through a water molecule. The H207E mutant has been shown to have higher affinity for iron because negative charges near Lys-206 are expected to stabilize the dilysine trigger favoring the iron-bound form (Yang et al., 2000). The $pK_a$ value found in this study is consistent with this suggestion.

The $pK_a$ of Glu-212, which is hydrogen-bonded to Ser-208, is strikingly low in all forms but particularly in the closed forms because the electrostatic potential near Glu-212 is mostly positive and so favors deprotonation. This positive potential is mainly due to the anion binding site that consists of Arg-124 and the N-terminus of the $\alpha$-helix number 5 (Baker, 1994) and which is located in the neighborhood of Glu-212. In Lf the residues equivalent to Glu-212, Lys-206, and Lys-296 are Glu-216, Arg-210, and Lys-301, respectively. Lf has no dilysine trigger, and Glu-216 hydrogen-bonds to Lys-301. As well as differences in the structure, it is possible that the electrostatic potential in $sTf_N$ favors formation of the dilysine interaction both directly, by promoting deprotonation of Lys-206, and indirectly, by stabilizing the appropriate conformation of Glu-212.

The total charges of various forms of the protein as a function of pH are shown in Fig. 1. The curves have the typical shape and the isoelectric point occurs at a pH of ~6.6 for the open apoform of $sTf_N$ and at slightly smaller values for the closed conformations. The experimental value for the complete sTf is in the range 5.4–5.9 (Brock, 1985) and so the agreement is reasonable, given that we do not take into account the whole protein. It also implies that both lobes display similar properties as far as the total charge and the isoelectric point are concerned. The number of protons that are released upon iron binding can be calculated to be ~2.6 at pH 7.5 or 3.6 if we assume that carbonate is released as bicarbonate. This is in agreement with the proposed value of 2.5–3 protons per binding site (Bates and Schlabach, 1975;

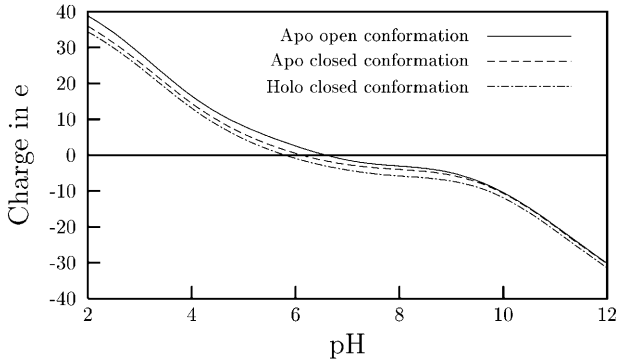


FIGURE 1 Total charge of various forms of $sTf_N$ as a function of pH.

Chasteen and Williams, 1981; Pakdaman and El Hage Chahine, 1996).

To see how the electrostatic potential changes with pH, we performed a range of calculations at different pH values using protonation states for the protein determined from the $pK_a$ calculations. The potential for the open apoform of $sTf_N$ is illustrated in Fig. 2, from which it can be seen that the potential in the iron-binding cleft is positive on average. The most positive part is located in the N2 domain at the end of helix 5 and near Arg-124 and the two tyrosines of the iron-binding site in the neighborhood of Lys-206. This region is close to the carbonate binding site. A second region, with a negative potential, is observed on the opposite side of the cleft, near Asp-63 and His-249 and coincides with part of the iron-binding site. The most widely accepted hypothesis for iron uptake says that carbonate binds first, followed by iron. To test the validity of this hypothesis, we performed a calculation on the open form of sTf with carbonate located at the anion binding site (it was placed by superimposing the N2 domain of the holoprotein with that of the apoprotein). The resulting surface (see Supplementary Material) shows that a low potential is created on both sides of the iron-binding cleft, which is in agreement with the fact that iron binding will be enhanced after the synergistic anion binds.

Fig. 3 displays an electrostatic potential surface of the holoform of $sTf_N$ at pH 5.3. The surface is sliced near the binding site and shows a cavity whose existence has been noted by several authors (Anderson et al., 1989; Hall et al., 2002) and which is conserved among members of the transferrin family, including the lactoferrins and ovotransferrins. The cavity is connected to the surface of the protein by two canals and has a mainly negative electrostatic potential even at low pH. This cavity-canal, which is occupied by water molecules and is terminated by Arg-124

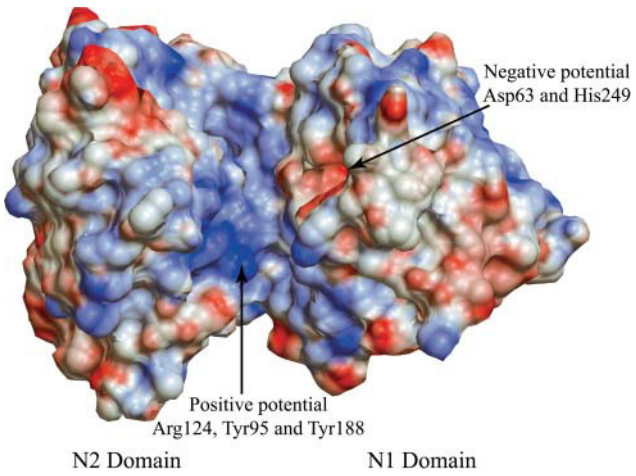


FIGURE 2 Electrostatic potential surface for the open apoform of $sTf_N$ at a pH of 7.4. Red corresponds to areas of negative potential ($<-7\ k_BT$), blue to positive potential ($>7\ k_BT$), and gray to intermediate potential.

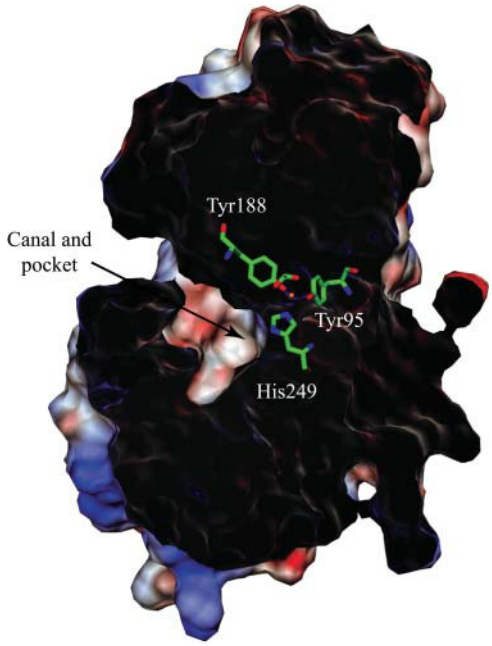


FIGURE 3 Electrostatic potential surface for the closed holoform of $sTf_N$ at a pH of 5.3, cut so as to expose His249.

and His-249, could therefore play an important role in iron release by attracting protons to the binding site.

An analysis of the electrostatic potential obtained by cutting the protein near the dilysine trigger clearly shows that the potential is positive behind the iron-binding site and near the hinge region (data not shown). This is probably related to the fact that this region can bind a nonsynergistic anion (He et al., 1999) which plays a role in iron release.

## Molecular dynamics simulations

Molecular dynamics simulations were performed for the open and closed apoforms of $sTf_N$ with protonation states appropriate to a pH of 5.3. The simulations consisted of 520 ps of equilibration followed by 1.5 ns of data collection. To assess the quality of the simulations, we monitored the potential energy, the root mean-square coordinate deviation (RMSCD) of trajectory structures compared to the starting structures, the surface area accessible to solvent, and the radius of gyration. We used relatively strict criteria to estimate whether the proteins were equilibrated because we wanted to ensure that we had a stable dynamics and a reasonable convergence to equilibrium before carrying out free-energy calculations.

For the open form, the potential energy changes rapidly during the first 20 ps of equilibration and then more slowly, until at ~500 ps, its average stabilizes. The closed form shows a quicker stabilization at ~300 ps. The RMSCDs for both forms along the trajectory are shown in Fig. 4. The mean RMSCD is higher for the open form, ~2.4 as opposed

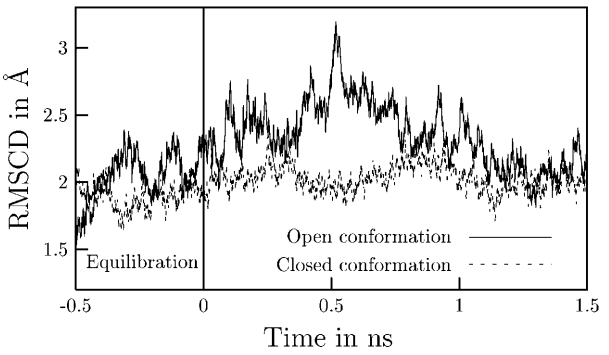


FIGURE 4 Root mean-square coordinate deviations with respect to the crystallographic structures of the open and closed apoforms of $sTf_N$ along the molecular dynamics trajectories. The RMSCD between the crystallographic open and closed forms is 6.8 Å for the complete N-lobe, 1.2 Å for the N1 domain, and 0.8 Å for the N2 domain.

to ~2.0 Å, and also fluctuates much more, attaining values as high as 3.2 Å and as low as 1.8 Å. Despite these large fluctuations, the integrity of the structure of the open form is clearly maintained as can be seen by observing other properties, such as its accessible surface area (see Fig. 5) and its radius of gyration throughout the dynamics (data not shown). The radii of gyration for both proteins seem to be stable, even during the equilibration phase, and have values of ~38.2 Å in both cases.

To further investigate how the geometries of the proteins change along the trajectory, we calculated RMSCDs between each of the structures along the 1.5-ns trajectory for the closed and open forms. The results are shown in Fig. 6. The upper-left-hand corner of the matrix gives the RMSCDs for the structures of the closed form and the lower-right-hand corner for the structures of the open form. Clearly, the RMSCDs for the open form are much higher than for the closed form. The highest values are ~3.7 Å and occur between structures toward the end of the simulation (1.25 and 1.45 ns) and from the beginning (0.2 and 0.5 ns). In between these two regions of high RMSCD, there is a region in which the RMSCD descends to the low value of 1.8 Å, reinforcing the idea that

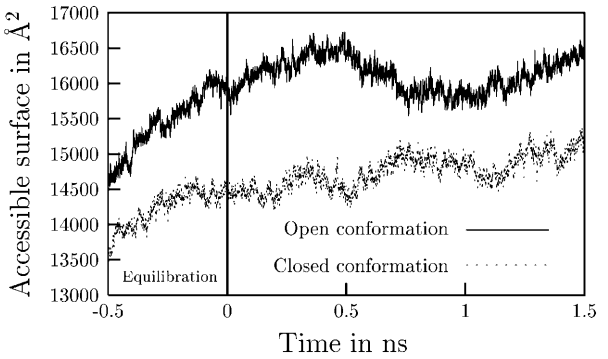


FIGURE 5 Accessible surface area of the open and closed apoforms of $sTf_N$ along the molecular dynamics trajectories.

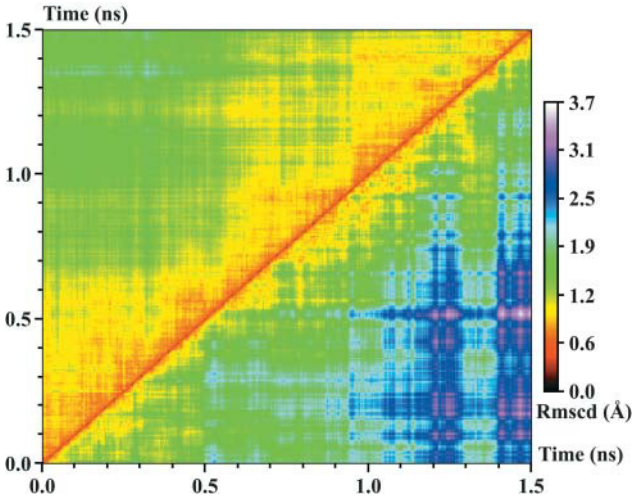


FIGURE 6 Matrix of root mean-square coordinate deviations between structures generated by the molecular dynamics simulations of the open and closed apoforms of $sTf_N$. (*Upper-left-hand corner*) Closed form and (*lower-right-hand corner*) open form. The RMSCDs are in Å.

the protein is not unfolding but suggesting that it is undergoing some sort of periodic motion. In contrast to the open form, the closed form shows RMSCDs that are much smaller, with the highest values being ~2.0 Å. These results indicate that the open form is much more mobile than the closed form. This is confirmed by inspection of the RMSCDs on the diagonal of the matrix between structures that are from similar times. For the closed form, there are large triangles of low RMSCD (*red* and *yellow*), indicating that the closed form oscillates in relatively restricted regions of structure space before jumping to configurations of higher RMSCD. The open form, however, rapidly jumps between structures of very different RMSCD.

A more detailed picture of the mobility of different regions of the protein can be obtained by calculating the Debye-Waller factors for the atoms from the simulation. We used the following formula to calculate these factors from the atomic fluctuations (Karplus and McCammon, 1983),

$$B = \frac{8\pi^2}{3}\langle \Delta r^2 \rangle_{\mathrm{dyn}}, \tag{8}$$

where $\langle \Delta r^2 \rangle_{\mathrm{dyn}}$ is a fluctuation in an atomic position. It should be noted that this formula ignores effects due to static disorder which are present in the crystal, and means that the baseline of the experimental B-factors will be higher than those determined from the simulations.

Figs. 7 and 8 display the backbone-atom Debye-Waller factors of the closed and open forms determined experimentally and from our simulations. Overall, there is reasonable agreement between theory and experiment regarding the most agitated parts of the closed form of the protein. Quantitatively, the agreement is less good (a correlation coefficient of 0.66 after removing 16 strongly deviating points out of 1312),

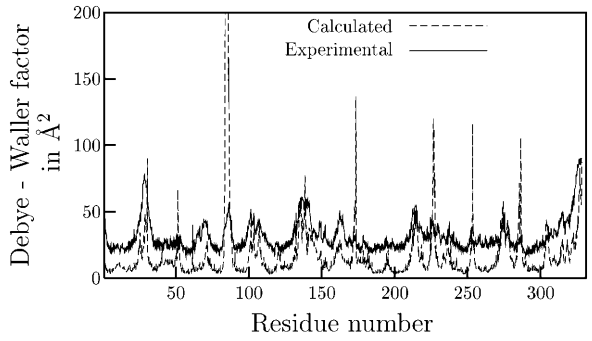


FIGURE 7 Debye-Waller factors for the backbone atoms of the closed form of $sTf_N$. The experimental values were taken from the crystallographic structure of the holoprotein and the theoretical values were determined from the simulation.

mainly because some mobile parts of the protein are much more mobile in the simulation than in the crystal. This is probably due to the fact that certain regions in the crystal are constrained by contacts with other protein molecules in the unit cell. Five parts of the protein have unusually high Debye-Waller factors. All of them are located on loops and are well-exposed to solvent. They include the sequence formed by the residues Ser-87, Lys-88, and Glu-89 and the isolated residues Cys-177, Asp-229, Met-256, and His-289.

Comparison between theory and experiment is less straightforward for the open form because the experimental structure was determined at a temperature of 113 K (Jeffrey et al., 1998). The mobile parts of the protein found in the x-ray structure are also seen in the simulation but the former have much smaller B-factors. Once again, the most mobile regions are loops which are accessible to solvent and include the sequences Ser-87–Glu-89 and Ser-287–Lys-291 and the residues Pro-31, Ala-54, Thr-181, Ala-215–Lys-217, and Asn-230. A comparison of the calculated B-factors for the open and closed forms shows that the open form has many more mobile regions.

It has been hypothesized that the open structure of sTf occasionally samples the closed state (Baker et al., 2002; Navati et al., 2003). To check this, we calculated how the angle between the two lobes varies along the trajectory, the results for which are shown in Figs. 9 and 10. To evaluate the angle variation, we first superimposed the N1 domain of each dynamical structure onto the N1 domain of the chosen crystallographic reference. The angle variation was then determined as the rotation angle necessary to superimpose the N2 domain of the dynamical structure onto the N2 domain of the same crystallographic structure. Although this way of measuring the angle is approximate, the variation is evidently much greater for the open form than for the closed form. Further analysis indicates that the interdomain angle for the open-form structures can decrease by as much as 20° but that the open-form structures remain substantially different from those of the closed form. Thus, our dynamics simulations suggest that the open form does not sample directly the closed state but rather intermediate states with some closed-form character.

A more detailed picture of the dynamics can be obtained by determining how the distances between critical residues in the active site change during the simulations. In the open form, the distance between the OH atom of Tyr-95 and the OD1 atom of Asp-63 is most noteworthy as it fluctuates between 4.3 and 11.6 Å with an average of 6.3 ± 1.5 Å. In the crystallographic structure its distance is 8.5 Å, which shows that the two residues can be quite close even without the iron and also attests to the flexibility of the open-form structure.

In the closed form, the distance between the lysines of the dilysine bridge is perhaps most interesting. In our simulation, both lysines were protonated and there was no iron in the binding site, so it might be expected that the two lysines repel each other, increasing the distance between them. This was indeed observed as the lysines adopt a conformation in which the distance between them is quite constant (6.7 ± 0.4 Å). In this conformation (see Supplementary Material), the main chain of His-249 changes conformation and moves toward the binding pocket whereas the side chain of Lys-296 takes

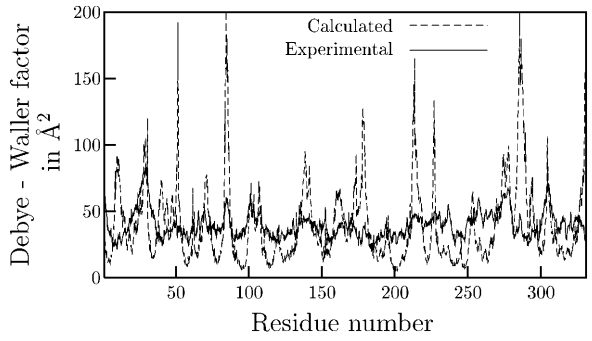


FIGURE 8 Debye-Waller factors for the backbone atoms of the open form of $sTf_N$. The experimental values were taken from the crystallographic structure of the apoprotein and the theoretical values were determined from the simulation.

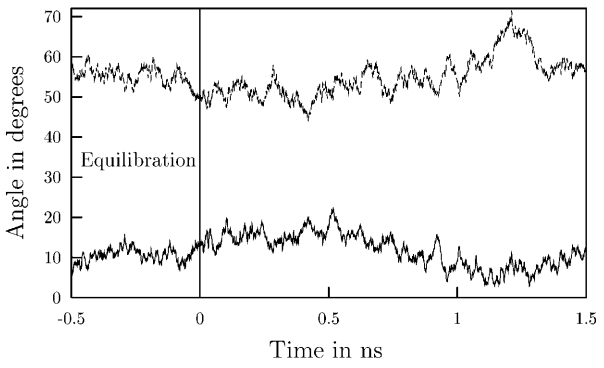


FIGURE 9 Variations of the angles between the N1 and N2 domains of $sTf_N$ for the open form. The variations were determined with respect to the crystallographic open (*solid line*) and crystallographic closed (*dashed line*) structures.

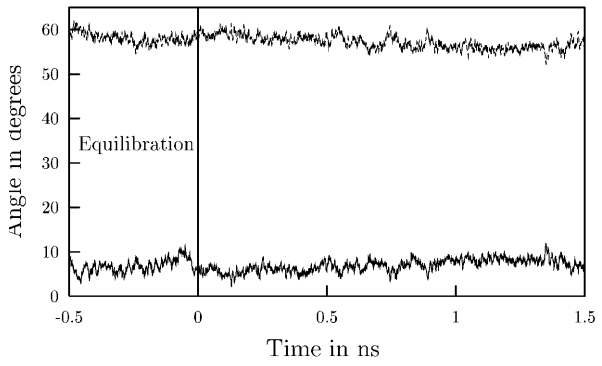


FIGURE 10 Variations of the angles between the N1 and N2 domains of $sTf_N$ for the closed form. The variations were determined with respect to the crystallographic closed (*solid line*) and crystallographic open (*dashed line*) structures.

the place of iron and forms a hydrogen bond with Asp-63. The latter also has hydrogen bonds with Tyr-188 and Arg-124. It seems therefore that the negative charge of the carboxylate of Asp-63 is neutralized by Lys-296, Arg-124, and Tyr-188.

Although both Lys-206 and Lys-296 were protonated in the simulations, the closed form appears quite stable and no domain opening was seen. To help analyze this stability, we investigated the hydrogen-bond network between the N1 and N2 domains (domains were defined as stated by MacGillivray et al., 1998). During the simulation, there are two very stable hydrogen bonds between the side chains of Asp-63 and Arg-124, each of which is present >90% of the time. These hydrogen bonds are not present in the crystallographic structure as Arg-124 is hydrogen-bonded to carbonate. Asp-63 also has a hydrogen bond with Tyr-188 which is present 70% of the time. Throughout the simulation, the average number of hydrogen bonds was $7 \pm 1.5$, to be compared with the ~25 hydrogen bonds present in the crystallographic structure. The simulation seems therefore to show that the direct hydrogen-bonding network between both domains has a relatively small influence on the stability of the closed conformation, although it should be emphasized that we did not study the usual closed conformation, which has carbonate and iron in the active site, and that there will be solvent molecules that bridge the domains.

## Essential dynamics analysis

The conventional analysis of the simulations strongly suggests that both forms of $sTf_N$ are equilibrated, but that the open form displays some reversible large-scale motions which would explain the great variation in the open-form RMSCDs and would be consistent with the notion that dynamics plays an important role in the function of transferrins (Baker et al., 2002). To further probe the results of the simulations, we have performed an essential dynamics analysis which is designed to identify important large-scale motions in dynamics trajectories.

The essential analysis was performed by taking into account all heavy atoms of the protein (2586 atoms or 7758 degrees of freedom) and using 3000 structures from the last 1.5 ns of the dynamics simulations. As expected, the covariance matrices had only 3000 nonzero eigenvalues. Fig. 11 shows the mean-square fluctuations generated by the eigenvectors of the covariance matrices plotted against vector index for the open and closed forms of $sTf_N$.

Comparison of the eigenvalues obtained for the open and closed form shows that the open form has mean fluctuations, on average, twice as high as the closed form, which implies that the open structure is more mobile. As reported for other simulations, there are only a few eigenvectors with large eigenvalues, indicating that the essential subspace which encapsulates the large-scale protein motions is small relative to the total number of degrees of freedom (Amadei et al., 1993; van Aalten et al., 1995; Cregut et al., 1998). In our case, the first five eigenvectors are sufficient to account for >70% of the total motion in the open form and >55% in the closed form (see Fig. 12). For both simulations, the first five eigenvectors induce movements of the C-terminal residues and the loops of the protein and, to a lesser extent, of the ends of some helices. In both forms, the cores of each domain, which are constituted of $\beta$-strands, are quite rigid whereas the external helices are more mobile. In the closed form, helix 11 is the most strongly involved of the helices in the motions corresponding to the first five essential modes. The mobility of the helices in the open form is much higher than for the closed form and the motions are, in general, much more widely spread throughout the protein.

For the open form, the first and the third essential modes induce a concerted motion of the helices of both domains. Projection of the trajectory structures onto these modes (see Figs. 13 and 14 and the films in the Supplementary Materials) shows that the first mode induces a concerted hinge-twist (Grossmann et al., 1998) motion between both domains in contrast to a shear mechanism (Gerstein et al., 1994, 1993).

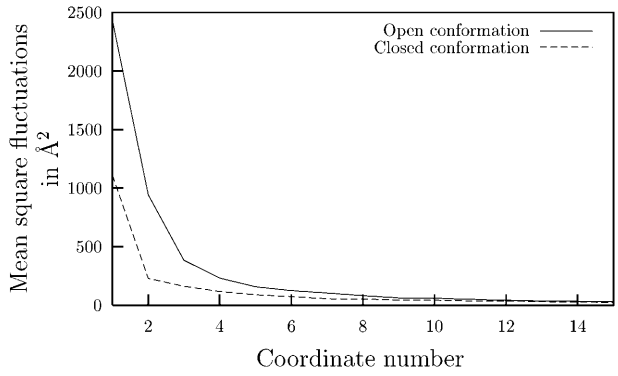


FIGURE 11 Mean-square fluctuations of the new coordinates defined by the essential dynamics analysis. The new coordinates are numbered in decreasing order of the size of their mean-square fluctuations and the first 15 coordinates are shown.

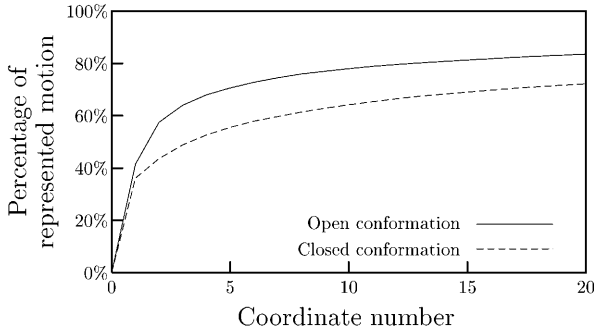


FIGURE 12 Percentage of total protein motion as a function of the number of essential mode coordinates for the simulations of the open and closed forms of $sTf_N$.

The direction of this axis and the rotation angle were calculated for all trajectory structures by first superimposing their N1 domains onto the N1 domain of the first structure of the dynamics and then superimposing the N2 domains. The axis of rotation was found to be very stable, with deviations of $<5\%$ between structures, whereas the angle fluctuates over a range of $>20°$ between $-6$ and 18°. Analysis of the motion induced by the first mode shows that it corresponds almost entirely to rotation about the hinge-twist axis as the correlation coefficient between the two motions is 0.999. In the motion, helix 11 is relatively still and the axis for the motion is located, to a good approximation, near the $\beta$-strands connecting the N1 and N2 domains. In particular, it passes close to the residues Val-205, Ala-203, Val-202, Val-98, and Ala-94, which are all hydrophobic and are located on two $\beta$-strands—one in the core of the N2 domain and the other just after the $\beta$-strands connecting the two domains.

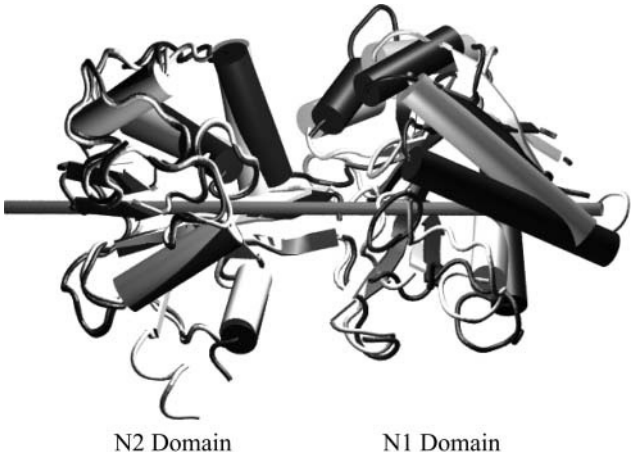


FIGURE 13 Motion induced by the first essential mode from the simulation of the open form of $sTf_N$. The image was generated by displacing the atoms of a representative open-form structure along the mode in positive and negative directions. The N2 domains of the resulting structures were then superimposed. Also shown is the axis corresponding to the hinge-twist motion.

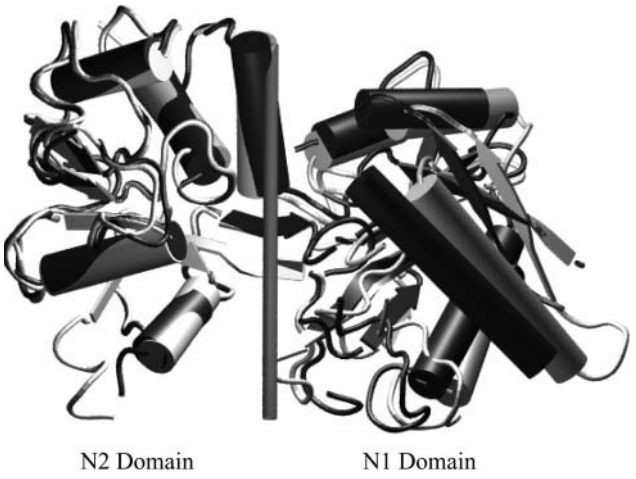


FIGURE 14 Motion induced by the third essential mode from the simulation of the open form of $sTf_N$. The image was generated by displacing the atoms of a representative open-form structure along the mode in positive and negative directions. The N2 domains of the resulting structures were then superimposed. Also shown is the axis corresponding to the hinge-bending motion.

The movement corresponding to the third eigenvector is a hinge-bending motion. The axis for the movement was found using the same method as above and it is almost orthogonal to the axis of the hinge-twist motion induced by the first mode and so also to the $\beta$-strand connecting the two domains. As for the first mode, the axis is found to be very stable (its direction deviates by $<2\%$) and the hinge-bending motion corresponds almost entirely to the motion induced by the third mode (the correlation coefficient is 0.987). The fluctuation of the hinge-bending angle between structures is also large, with a range of 15° and minimum and maximum values of $-5$ and 11°, respectively. The residues that move most during the hinge-bending motion are the C-terminal residue, Asp-337, followed by Lys-88 and Ser-87. This is important, inasmuch as the latter two residues were seen to have extremely high mobility in the dynamics of the open form, implying that this mobility is, in large part, due to the hinge-bending motion.

Visual inspection of the motions induced by the first and third essential modes of the open form, along with a more detailed analysis of the motions induced by the modes, shows that the hinge-twist and hinge-bending motions are almost sufficient to describe domain closure. This can be shown, for example, by superimposing structures of the open and closed forms and then performing the necessary rotations about the relevant axes. The region critical for these motions is the loop formed by residues 86–92 that lies between the **d** and **e** $\beta$-strands (Haridas et al., 1995).

## Free-energy calculations

The essential dynamics analysis confirms the higher mobility of the open structure but gives no indication of the relative

thermodynamic stability of both forms. Experimental results tend to show that the open apoform is more stable than the closed apoform (Mecklenburg et al., 1997; Grossmann et al., 1998). Likewise, no closed apoforms have been observed for sTf although closed apoforms have been observed for both lobes of mare Lf (Sharma et al., 1999) and for the C-lobe of human Lf (Baker et al., 1991). These results suggest that the closed apoform might be more stable in the case of Lf than in the case of sTf which could explain, in part, the better stability of iron-loaded Lf compared to iron-loaded sTf. To address this problem, we used the MM/PBSA method which, despite significant approximations, seems to be able to reproduce free-energy differences in respectable agreement with experiment, and with much lower computational effort for systems that differ substantially in structure (Kollman et al., 2000) than alternative simulation methods for determining free-energies.

The trajectories of both simulations were post-processed with the MM/PBSA method, using structures collected at 5-ps intervals. The main results of our calculations are given in Table 2 for the reaction going from the closed to the open form. Clearly the open form is thermodynamically more stable. Two, roughly equal, contributions make up the calculated energy differences. First, the open form has a more favorable electrostatic solvation energy, $\Delta\overline{G}_{\mathrm{solv-pol}}$, which more than makes up for its lower internal energy, $\Delta E_{\mathrm{MM}}$, with respect to the closed form. The internal energy is lower mainly due to the loss of nonbonding interactions between the N1 and N2 domains. Second, the entropy contribution favors the more flexible open form.

The energy calculations were performed using different force-field parameter sets and also different environmental ionic strengths, but the results do not seem to depend critically on either factor, given the overall magnitude of each term. It is evident that there is a great deal of statistical uncertainty in the energies. These are, in large part, due to the high flexibility of the open form of the protein because the hinge motions induce large structural movements that lead to large changes in the nonbonding energies between the N1 and N2 domains and in the solvation energy in the iron-binding cleft. The value of the entropy contribution to the free-energy difference between the open and closed forms ($\sim$60 kcal $\times$ mol$^{-1}$) is large but is similar in size to that obtained by van Gunsteren and co-workers for two possible conformers of the molten globule of $\alpha$-lactalbumin (Schafer et al., 2000, 2002).

In Table 2, we do not state an uncertainty in the entropy values because the entropy is a global property that does not depend upon a single structure. Nevertheless, it is important to be able to assess whether our entropy values have converged. To do this we calculated the entropies for the open and closed forms as a function of time (or number of structures), and also the contributions of various numbers of modes to the entropy difference between the two forms. The results are shown in Figs. 15 and 16, respectively. It is clear from Fig. 15 that the entropies for the two forms are converging but have not *yet* converged, whereas Fig. 16 shows that the entropies for each batch of modes converge quite rapidly. Thus, the lack of convergence is mainly due to the fact that, as the length of the trajectory increases, the number of nonzero modes contributing to the entropy increases. To include all such modes in our calculation we would need a minimum of 7758 dynamical structures—which would imply a simulation of at least 4 ns, a length that is currently beyond our computational resources. Although not fully converged, it is evident that the entropy plays an important and even the dominant role in the free-energy difference between the open and closed forms.

## DISCUSSION

### The dilysine trigger

The most interesting results of our $pK_a$ calculations concern the dilysine bridge and show that Lys-206 has a low $pK_a$ value whereas Lys-296 has a high $pK_a$ value. This is consistent with the commonly accepted view, which holds that at high and neutral pH values one lysine is deprotonated and hydrogen-bonded to the other lysine. This forms a bridge between the two domains, thereby strengthening the closed conformation and favoring iron-binding. At low pH both lysines would be protonated, and the repulsion between them would trigger the opening of the cleft and iron release.

This view is not, however, consistent with the results of our dynamics simulations of the closed form in which we saw that the two lysines have enough freedom to locally change conformation without triggering domain opening. This result is also supported by some experimental results. Thus, in the sTf H249E mutant (MacGillivray et al., 2000),

**TABLE 2 The MM-PBSA free-energy differences between the open and closed apoforms of human $sTf_N$**

| Method | $\Delta\overline{E}_{\mathrm{MM}}$ | $\Delta\overline{G}_{\mathrm{solv-pol}}$ | $\Delta\overline{G}_{\mathrm{solv-non-pol}}$ | $\Delta U$ | $T\Delta S$ | $\Delta\overline{G}$ |
|---|---|---|---|---|---|---|
| CHARMM (150) | 389 $\pm$ 227 | $-428 \pm 186$ | 7 $\pm$ 3 | $-32 \pm 112$ | 59.7 | $-92.1$ |
| CHARMM (0) | 389 $\pm$ 227 | $-427 \pm 186$ | 7 $\pm$ 3 | $-31 \pm 112$ | 59.7 | $-90.7$ |
| OPLS-AA (150) | 381 $\pm$ 252 | $-429 \pm 184$ | 7 $\pm$ 3 | $-41 \pm 149$ | 59.7 | $-100.4$ |
| PARSE (150) | 389 $\pm$ 227 | $-452 \pm 206$ | 7 $\pm$ 3 | $-57 \pm 129$ | 59.7 | $-116.6$ |

The definitions of the terms follow those described in the text (see Eq. 2) except for $\Delta U$ which is the sum of the three previous columns in the table. CHARMM, OPLS-AA, and PARSE refer to the force-field parameter sets used for the calculations whereas the brackets after these names give the ionic strengths (in mM) used in the solution of the Poisson-Boltzmann equation. All energies are in kcal $\times$ mol$^{-1}$.

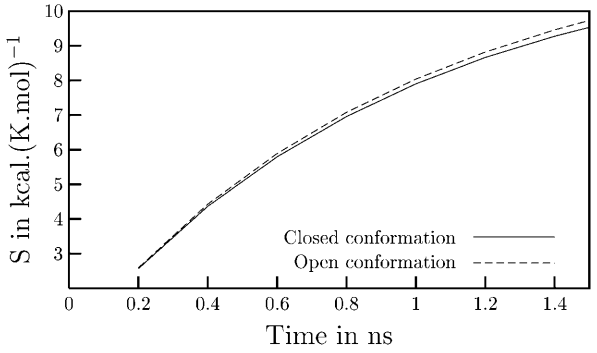


FIGURE 15 The entropies ($S$) for the open and the closed conformations of $sTf_N$ as a function of simulation time.

Glu-249 has nearly the same conformation as His-249 does in wild-type sTf, but the dilysine trigger is abolished and Lys-296 undergoes a conformational change related to the one found in our dynamics simulation of the closed form. Despite the disruption of the dilysine trigger, the mutant is found to be only slightly more acid-stable than the wild-type. In addition, bovine Lf has two lysine residues that are equivalent to those in the dilysine trigger, but these residues are not hydrogen-bonded—probably due to electrostatic effects, as there is no obvious steric hindrance preventing their interaction. To summarize, we think that the dilysine trigger clearly has an influence on iron release, but we agree with Lindley (2001) that it is unlikely that the protonation of one of the lysines of the dilysine bridge results in domain opening.

An alternative mechanism for domain opening that we propose here is based, in part, on the fact that Lys-296 is protonated at high pH values. This lysine is close to the iron-binding residue Tyr-188 (3.17 Å) and so can be hydrogen-bonded to and maybe even transfer its proton to the oxygen of the tyrosyl group. At high pH, Lys-296 is protonated whereas Lys-206 is unprotonated and hydrogen-bonded to Lys-296. Thus, the formation of a hydrogen bond between

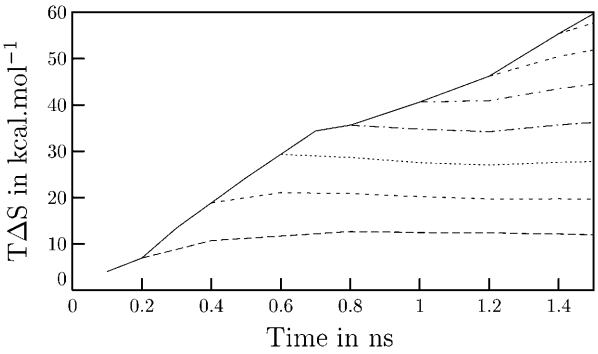


FIGURE 16 The entropy difference between the open and the closed conformations of $sTf_N$ as a function of simulation time and as a function of the number of modes used to calculate the entropies. The dotted and dashed lines correspond, from the bottom up, to the entropy differences calculated with 400, 900, 1200, 1600, 2000, 2400, and 2800 quasiharmonic modes.

Lys-296 and Tyr-188 and the transfer of the proton is not favored. As the pH decreases, Lys-206 becomes protonated and the resulting electrostatic repulsion forces the transfer of one proton of Lys-296 to Tyr-188, thereby weakening iron binding. It is possible that protonation and proton transfer could occur in a concerted fashion. This reaction would only require very minor conformational changes in the protein and could benefit from dynamical fluctuations in its structure. Indeed, Tyr-188 and Lys-206 belong to the N2 domain whereas Lys-296 belongs to the N1 domain. Therefore, hinge-bending fluctuations of the domains in the closed form will greatly support the transfer of a hydrogen from Lys-296 to Tyr-188, as this motion tends to decrease/increase the distance between Lys-296–Lys-206 and Lys-296–Tyr-188 in a concerted way, and thus destabilize the protonation state of Lys-296 and favor the transfer of a proton to Tyr-188 in a concerted manner. It is also possible that the proton transfer could be enhanced by the tunnel effect, which is known to be important in certain enzyme reactions (Antoniou et al., 2002; Sutcliffe and Scrutton, 2002; Knapp and Klinman, 2002).

This mechanism has some experimental support. First, Tyr-188 is critical for iron binding (He et al., 1997; Kuser et al., 2002) as its mutation into phenylalanine results in the loss of iron binding and so its protonation could therefore trigger iron release. Second, the mechanism is consistent with the pH stability of the C-lobe compared with the N-lobe and with the behaviors of various mutations of the N-lobe. In the C-lobe, the dilysine bridge is not observed as Arg-641 (porcine sTf numbering; Hall et al., 2002) replaces Lys-296. Arginine has a $pK_a$ two units higher than lysine, and so it will be harder to protonate the tyrosine ligand of iron although it may be possible at low pH values. The proton to be transferred to Arg-641 could come from the lysine equivalent to Lys-206 or, alternatively, from a water molecule (or a hydroxonium ion) which is hydrogen-bonded to Arg-641 in the crystal structure. Equally, in the N-lobe, in mutations of the dilysine residues (Steinlein et al., 1998; Nurizzo et al., 2001) or in mutations that result in a conformational change of these lysines (MacGillivray et al., 2000), a water molecule often takes the place of the NZ atom of the missing lysine in the hydrogen-bonding network. The presence of water would still permit the shunting of protons but would make it more difficult and so favor iron release at lower pH values. These arguments apply to mutants in which alanine replaces lysine (Nurizzo et al., 2001) as well as to glutamate and glutamine mutants (He et al., 1999; Yang et al., 2000). Mutants of the dilysine trigger also release iron more slowly than the wild-type, as the disruption of the trigger and its replacement by a water molecule will not allow the same rapid, concerted proton shunt from the lysines to Tyr-188.

The explanations in the previous paragraph also apply to the situation in Lfs and some of its mutants which do not have dilysine bridges and release iron at lower pH than sTf. In $Lf_N$, an arginine, Arg-210, replaces the equivalent of $sTf_N$

Lys-206 in the sequence but spatially occupies the place of Lys-296 whereas there is a water molecule in the place of Lys-206 (Haridas et al., 1995). This is reminiscent of $sTf_C$ discussed in the previous paragraph. Mutation of Arg-210 to lysine does destabilize iron-binding somewhat but not as much as would be expected if a dilysine trigger was formed. Indeed, the crystallographic structure shows that the lysines do not interact (Peterson et al., 2000). Instead, a water molecule occupies the place occupied by the NZ atom of Lys-296 in $sTf_N$ whereas Lys-210 is in the same position as $sTf_N$ Lys-206.

Two other points seem to be critical for the mechanism that we propose above. They are the electrostatic potential and the hydrogen-bond network in the region of the dilysine trigger. As pointed out by our electrostatic calculations, the potential around Lys-206 and Lys-296 is clearly positive. This is consistent with experimental results (He et al., 1999; Baker, 1994) showing that this region can bind a nonsynergistic anion which is essential for iron release. As this anion is probably next to Lys-296 (He et al., 1999), it will prevent proton transfer from Lys-296 to Tyr-188 at high pH, which seems to be the case experimentally. At low pH, the rate-determining step for iron release is not protonation of Tyr-188 but the arrival of a chelator replacing Tyr-188. The anion could therefore play this role at least temporarily until the domain has opened.

The second point deals with the hydrogen-bond network around Lys-296. In the $pK_a$ calculations, Glu-83 and Tyr-85 have very low and very high $pK_a$ values, respectively, in both open and closed forms, because they are hydrogen-bonded. However, in the closed conformation, Tyr-85 is only 3.3 Å from Lys-296 and is protonated. Therefore, there is a hydrogen bond between the tyrosine oxygen and Lys-296, and so Tyr-85 could modulate the dilysine trigger mechanism through this hydrogen bond. Experimentally, the mutant Tyr-85/Phe-85 (He et al., 1998) releases iron at higher pH than the wild-type, which is compatible with our mechanism, inasmuch as the hydrogen bond between Tyr-85 and Lys-296 will hinder proton transfer from Lys-296 to Tyr-188 because it stabilizes the protonated Lys-296. Thus, disruption of this hydrogen bond in the mutant leads to iron release at higher pH than the wild-type. Interestingly, Glu-83 is also hydrogen-bonded to a water molecule that bridges with His-249. Overall, it appears that this hydrogen-bond network could act as a means of communication between the dilysine trigger, Tyr-188, and His-249, which helps to coordinate domain opening, disruption of iron's coordinate shell for iron release, and His-249 conformational change.

## His-249 and iron release

As discussed in the Results section, His-249 is located at the back of a pocket that is linked to the surface of the protein through a canal which has a negative potential even at low pH. It is possible, therefore, that this pocket provides a means of attracting protons (or hydroxonium ions) for the protonation of His-249, triggering its conformational change (observed in the simulation of the closed form) and leading to its displacement from the coordination shell of iron. Given the conformation of His-249 in the pocket, it is the atom ND1 which is accessible for protonation. For this protonation to occur, it would mean that His-249 is not in its neutral state when bound to iron, as commonly accepted (Lindley, 2001), but in its doubly deprotonated, imidazolate form which has been proposed to occur for other metal-bound histidines (see El Yazal and Pang, 1999, and references therein). The $pK_a$ value for the imidazole/imidazolate pair is ~14 under standard conditions but this will certainly be reduced in the presence of a highly charged ion such as $Fe^{III}$. In our current work on the iron-bound form of $sTf_N$, we have performed extensive quantum chemical calculations, which do indeed support the hypothesis of an imidazolate ligand for the iron.

Glu-83 is likely important in conformational changes of His-249, as it hydrogen-bonds to the ND1 atom of this residue via a bridging water molecule, and experiments in which Glu-83 is mutated to alanine have been interpreted as showing that the iron loses the His-249 ligand (He et al., 1998). This could be because removal of Glu-83 makes the protonation of His-249 easier and so causes a weakening of the interaction between it and the iron. Addition of an anion to the E83A mutant restores the coordination of the iron presumably by mimicking the effect of Glu-83.

It is clear that second-shell residue effects appear to be essential for understanding the iron-release mechanism. The two steps of iron release proposed here take into account these effects and we hypothesize that they are interdependent. Indeed, the hydrogen-bond network His-249–water–Glu-83–Tyr-85–Lys-296 could be a means of removing His-249 and Tyr-188 from the iron coordination shell in a concerted way.

The other step involved in iron release concerns the protonation of the synergistic anion, carbonate. Our calculations do not give any information for this step, inasmuch as the $pK_a$ of the carbonate was not calculated. However, carbonate is hydrogen-bonded to Arg-124 and this latter residue lies at the back of the same cavity as His-249. This implies that protons coming into the cavity could lead to the protonation of carbonate, via Arg-124, in addition to His-249.

The precision of our calculations does not allow us to propose an order for the three steps of iron release. We therefore take the commonly accepted order (Lindley, 2001): 1), protonation of carbonate leading to arginine conformational change; 2), protonation of His-249 leading to a weakened binding of iron and a possible conformational change in Glu-83, which could trigger the third step through the hydrogen-bond network; and 3), protonation of Lys-206 and proton transfer to Tyr-188 via Lys-296. Binding of a nonsynergistic anion probably helps in both domain opening and removal of Asp-3.

### Iron binding

Our electrostatic potential calculations have shown that the region near Tyr-95 and Tyr-188 at the surface of the N2 domain is attractive for anions in the open apoform and that, after carbonate binding, this region has a negative electrostatic potential in the region surrounding Asp-63 and His-249 on the N1 domain. These calculations support the currently favored mechanism that states that carbonate binds first to the N2 domain followed by iron-binding to Tyr-95, Tyr-188, and carbonate. Domain closure then occurs allowing Asp-63 and His-249 of the N1 domain to bind iron. In the crystallographic structure, Asp-63 is located at ~9 Å from Tyr-95 (Baker et al., 2002; Mizutani et al., 2001), and so it is difficult to explain how electrostatic interactions between iron and Asp-63 alone are able to trigger domain closure. This led Hirose (Mizutani et al., 2001) to propose an alternative mechanism, involving the region around Lys-88, whereas other authors have invoked the ability of the protein to sample the closed or partially closed conformations (Baker et al., 2002; Navati et al., 2003). The results of our simulations show that despite the fact that the open structure is always far from the closed conformation, the distance between Asp-63 and Tyr-95 varies over a wide range—from as low as 4.3 to as high as 11.6 Å. This variability is due to movement of the domains, as well as of the relevant residue side chains, and supports the hypothesis that the dynamics of the protein is critically important for iron binding (Baker et al., 2002, 2003). These results also show that a sampling of closed-form conformations is not essential for there to be a short distance between Asp-63 and the iron. This is clearly consistent with solution x-ray scattering measurements that indicate that the open form is not continuously sampling the closed conformation in the absence of iron (Grossmann et al., 1998), and the hypothesis that Asp-63 is the residue that triggers domain closure (Grossmann et al., 1993).

Our simulations of the open apoform of the protein showed that $sTf_N$ undergoes significant, large-scale domain movements even in the absence of iron and at a low pH. These movements can be described as a combination of two independent motions, one a hinge-twist and another a hinge-bending. The former mode has the larger amplitude in our analysis, indicating that the hinge-twist would be the preponderant motion. Both of these movements have been proposed before. The hinge-bending was deduced from the crystallographic structures of the apo- and holoforms of the protein whereas the hinge-twisting motion was only evoked very recently (Grossmann et al., 1998). Our results also support the homology between transferrin and maltodextrin-binding protein in which domain closure has also been described as resulting from a combination of a hinge-bending and a hinge-twist (Sharff et al., 1992). The occurrence of two independent motions that cause domain closure could help to explain the ability of transferrins to bind different metals as it would give the protein extra flexibility in adapting its mechanism of closure to the size of the ligand.

Finally, it has been postulated for oTf that the region around Lys-88 plays an important role in domain closure (Mizutani et al., 2001). It is clear from our simulations that this region also has a high mobility in sTf, induced in part by the hinge-twist and hinge-bending motions that lead to domain closure, but we think it unlikely that this loop region is itself the driving force behind closure. Instead we postulate that this region and its high mobility could play a role in the recognition of iron-loaded sTf by the transferrin receptor (TfR). Binding of TfR to this region could also be a way in which TfR helps iron release. This hypothesis seems to be consistent with the Fe-Tf/TfR complex structure proposed by Harrison (Lawrence et al., 1999). Equally this structural segment is present in the N-lobes of Lf and oTf but is different in the C-lobes of sTf, Lf, and oTf, and so it could play a role in the difference observed between the iron-release properties of the two lobes when bound to TfR (Bali et al., 1991; Bali and Aisen, 1991).

## CONCLUSION

In this molecular modeling study, we have sought to gain further insight into the electrostatic and dynamical properties of the open and closed conformations of $sTf_N$ and the implications that these have for its iron-binding and iron-release mechanisms. Some of our findings are:

- Our $pK_a$ calculations support the fact that Lys-296 is protonated but that Lys-206 is deprotonated at neutral pH in the closed form of $sTf_N$. The molecular dynamics simulations show, however, that protonation of Lys-206 is not, in itself, sufficient to trigger domain opening and we propose a mechanism in which the dilysine bridge acts to protonate Tyr-188 instead. In addition, consideration of His-249 and the cavity-canal going to it, leads us to suggest that His-249 is not neutral when bound to iron but in its imidazolate form. This would mean that protonation of His-249, as well as that of carbonate, Lys-206, and Tyr-188, would be important for iron release.
- The simulation of the open form of the apoprotein demonstrates that this form is very mobile and experiences two important motions—a hinge-bending and a hinge-twist. This feature is essential for understanding iron-binding because these motions enable $sTf_N$ to sample conformations intermediate between the closed and the open crystallographic forms. It also demonstrates that Asp-63 could act as a domain closure residue because this residue can approach the iron-binding site closely as the position of the domains fluctuates.
- The sequence of residues Tyr-85–Thr-93 is very mobile in our dynamics simulations, a mobility which is

induced in large part by the motions leading to domain opening and closing of $sTf_N$. The striking correlation of the motions of these residues with the hinge-twist domain motion demonstrates the door-closure-like behavior of this loop. This, together with differences in the structure of this loop between the open and closed forms, leads us to propose that the loop could play a role in the recognition of iron-loaded $sTf_N$ by TfR.

At pH 5.3, the open apoform of $sTf_N$ is more stable than the closed apoform, which is consistent with the study of Grossmann et al. (1998). The difference in free energy between the two forms is large, and at least half of this is due to the higher entropy of the open form. Although there are large statistical uncertainties in the energy values that we calculate, we think that the difference that we obtain is qualitatively correct, as at pH 5.3 sTf releases iron—implying that the open form is more stable even when iron is present.

The studies that we report here concern, for the most part, $sTf_N$ with neither iron nor carbonate bound. Obviously, it is important to include these groups for a fuller understanding of the properties of $sTf_N$. Therefore, we are currently investigating in detail various models of the iron-binding site with quantum chemical methods and aim to continue with simulations of iron-loaded $sTf_N$ once these latter studies are complete.

## SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors thank the Institut de Biologie Structurale—Jean-Pierre Ebel, the Commissariat à l'Energie Atomique and the Centre National de Recherche Scientifique for support of this work.

## REFERENCES

Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Amadei, A., A. Linssen, and H. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.

Anderson, B., H. Baker, G. Norris, D. Rice, and E. Baker. 1989. Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 A resolution. *J. Mol. Biol.* 209:711–734.

Andricioaei, I., and M. Karplus. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289–6292.

Antoniou, D., S. Caratzoulas, C. Kalyanaraman, J. Mincer, and S. Schwartz. 2002. Barrier passage and protein dynamics in enzymatically catalyzed reactions. *Eur. J. Biochem.* 269:3103–3112.

Antosiewicz, J., J. McCammon, and M. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Baker, E. 1994. Structure and reactivity of transferrins. *Adv. Inorg. Chem.* 41:389–463.

Baker, E., B. Anderson, H. Baker, M. Haridas, G. Jameson, G. Norris, S. Rumball, and C. Smith. 1991. Structure, function and flexibility of human lactoferrin. *Int. J. Biol. Macromol.* 13:122–129.

Baker, E., H. Baker, and R. Kidd. 2002. Lactoferrin and transferrin: functional variations on a common structural framework. *Biochem. Cell Biol.* 80:27–34.

Baker, E., and P. Lindley. 1992. New perspectives on the structure and function of transferrins. *J. Inorg. Biochem.* 47:147–160.

Baker, H., B. Anderson, and E. Baker. 2003. *Bioinorganic Chemistry Special Feature*. Dealing with iron: common structural principles in proteins that transport iron and heme. *Proc. Natl. Acad. Sci. USA.* 100:3579–3583.

Bali, K., and P. Aisen. 1991. Receptor-modulated iron release from transferrin: differential effects on N- and C-terminal sites. *Biochemistry.* 30:9947–9952.

Bali, P., O. Zak, and P. Aisen. 1991. A new role for the transferrin receptor in the release of iron from transferrin. *Biochemistry.* 30:324–328.

Bates, G., and M. Schlabach. 1975. The reaction of ferric salts with transferrin. *J. Biol. Chem.* 248:3228–3232.

Briggs, J., J. Madura, M. Davis, M. Gilson, J. Antosiewicz, B. A. Luty, R. Wade, B. Bagheri, A. Ilin, R. Tan, and J. McCammon. 1989. University of Houston Brownian Dynamics Program User's Guide and Programmer's Manual, Rel. 5.1. University of Houston. Houston, TX.

Brock, J. 1985. Transferrins. *Metalloproteins.* 2:183–262.

Brock, J. 2002. The physiology of lactoferrin. *Biochem. Cell Biol.* 80:1–6.

Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimisation, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Brooks, B., D. Janezic, and M. Karplus. 1995. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.* 16:1522–1542.

Chasteen, N., and J. Williams. 1981. The influence of pH on the equilibrium distribution between the metal binding sites of human transferrin. *Biochem. J.* 193:717–727.

Cregut, D., G. Drin, J. Liautard, and L. Chiche. 1998. Hinge-bending motions in annexins: molecular dynamics and essential dynamics of apo-annexin V and of calcium bound annexin V and I. *Protein Eng.* 11:891–900.

David, L. 1996. Modélisation des interactions électrostatiques des biomoloécules en solution. Université Joseph Fourier-Grenoble I. France. PhD thesis.

Dewan, J., B. Mikami, M. Hirose, and J. Sacchettini. 1993. Structural evidence for a pH-sensitive dilysine trigger in the hen ovotransferrin N-lobe: implications for transferrin iron release. *Biochemistry.* 32: 11963–11968.

El Yazal, J., and Y. Pang. 1999. Ab initio calculations of proton dissociation energies of zinc ligands: hypothesis of imidazolate as zinc ligand in proteins. *J. Phys. Chem. B.* 103:8773–8779.

Feeney, D., and D. Komatsu. 1966. The transferrins. *Struct. Bonding.* 1:149–206.

Field, M. 1999. A Practical Introduction to the Simulation of Molecular Systems. Cambridge University Press, Cambridge, UK.

Field, M., M. Albe, C. Bret, F. Proust-De Martin, and A. Thomas. 2000. The DYNAMO library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* 21:1088–1100.

Gerstein, M., B. Anderson, G. Norris, E. Baker, A. Lesk, and C. Chothia. 1993. Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.* 234:357–372.

Gerstein, M., A. Lesk, and C. Chothia. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry.* 33:6739–6749.

Gilson, M. 1993. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins.* 15:266–282.

Gorinsky, B., C. Horsburgh, P. Lindley, D. Moss, M. Parkar, and J. Watson. 1979. Evidence for the bi-lobal nature of diferric rabbit plasma transferrin. *Nature*. 281:157–158.

Grossmann, J., J. Crawley, R. Strange, K. Patel, L. Murphy, M. Neu, R. Evans, and S. Hasnain. 1998. The nature of ligand-induced conformational change in transferrin in solution. An investigation using x-ray scattering, XAFS and site-directed mutants. *J. Mol. Biol.* 279:461–472.

Grossmann, J., A. Mason, R. Woodworth, M. Neu, P. Lindley, and S. Hasnain. 1993. Asp ligand provides the trigger for closure of transferrin molecules. Direct evidence from x-ray scattering studies of site-specific mutants of the N-terminal half-molecule of human transferrin. *J. Mol. Biol.* 231:554–558.

Hall, D., J. Hadden, G. Leonard, S. Bailey, M. Neu, M. Winn, and P. Lindley. 2002. The crystal and molecular structures of diferric porcine and rabbit serum transferrins at resolutions of 2.15 and 2.60 A, respectively. *Acta Crystallogr. D Biol. Crystallogr.* 58:70–80.

Haridas, M., B. Anderson, and E. Baker. 1995. Structure of human differic lactoferrin refined at 2.2 Å resolution. *Acta Crystallogr.* D51:629–646.

He, Q., A. Mason, B. Tam, R. MacGillivray, and R. Woodworth. 1999. Dual role of Lys-206-Lys-296 interaction in human transferrin N-lobe: iron-release trigger and anion-binding site. *Biochemistry.* 38:9704–9711.

He, Q., A. Mason, R. Woodworth, B. Tam, R. MacGillivray, J. Grady, and N. Chasteen. 1997. Inequivalence of the two tyrosine ligands in the N-lobe of human serum transferrin. *Biochemistry.* 36:14853–14860.

He, Q., A. Mason, R. Woodworth, B. Tam, R. MacGillivray, J. Grady, and N. Chasteen. 1998. Mutations at nonliganding residues Tyr-85 and Glu-83 in the N-lobe of human serum transferrin. Functional second shell effects. *J. Biol. Chem.* 273:17018–17024.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:27–28;33–38.

Jeffrey, P., M. Bewley, R. MacGillivray, A. Mason, R. Woodworth, and E. Baker. 1998. Ligand-induced conformational change in transferrins: crystal structure of the open form of the N-terminal half-molecule of human transferrin. *Biochemistry.* 37:13978–13986.

Jorgensen, W., D. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118:11225–11236.

Karplus, M., and J. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.

Karplus, M., and J. McCammon. 1983. Dynamics of proteins: elements and function. *Annu. Rev. Biochem.* 53:263–300.

Knapp, M., and J. Klinman. 2002. Environmentally coupled hydrogen tunneling. Linking catalysis to dynamics. *Eur. J. Biochem.* 269:3113–3121.

Kollman, P., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. Case, and T. Cheatham 3rd. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

Krzanowski, W. 1988. Principles of Multivariate Analysis. Oxford University Press, Oxford, UK.

Kuser, P., D. R. Hall, M. Haw, M. Neu, R. Evans, and P. Lindley. 2002. The mechanism of iron uptake by transferrins: the x-ray structures of the 18 kDa NII domain fragment of duck ovotransferrin and its nitrilotriacetate complex. *Acta Crystallogr. D.* D58:777–783.

Lawrence, C., S. Ray, M. Babyonyshev, R. Galluser, D. Borhani, and S. Harrison. 1999. Crystal structure of the ectodomain of human transferrin receptor. *Science.* 286:779–782.

Lee, D., and J. Goodfellow. 1998. The pH-induced release of iron from transferrin investigated with a continuum electrostatic model. *Biophys. J.* 74:2747–2759.

Levy, R., M. Karplus, and J. Kushick. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an $\alpha$-helix. *Macromolecules.* 17:1370–1374.

Lindley, P. 2001. Transferrins. *In* Handbook of Metalloproteins, Vol. 2. A. Messerschmidt, R. Huber, T. Poulos, and K. Wieghardt, editors. John Wiley & Sons, Chichester, UK. 793–811.

MacGillivray, R., M. Bewley, C. Smith, Q. He, A. Mason, R. Woodworth, and E. Baker. 2000. Mutation of the iron ligand His-249 to Glu in the N-lobe of human transferrin abolishes the dilysine ''trigger'' but does not significantly affect iron release. *Biochemistry.* 39:1211–1216.

MacGillivray, R., and K. Brew. 1975. Transferrin: internal homology in the amino acid sequence. *Science.* 190:1306–1307.

MacGillivray, R., S. Moore, J. Chen, B. Anderson, H. Baker, Y. Luo, M. Bewley, C. Smith, M. Murphy, Y. Wang, A. Mason, R. Woodworth, G. Brayer, and E. Baker. 1998. Two high-resolution crystal structures of the recombinant N-lobe of human transferrin reveal a structural change implicated in iron release. *Biochemistry.* 37:7919–7928.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

Mecklenburg, S., R. Donohoe, and G. Olah. 1997. Tertiary structural changes and iron release from human serum transferrin. *J. Mol. Biol.* 270:739–750.

Mizutani, K., B. Mikami, and M. Hirose. 2001. Domain closure mechanism in transferrins: new viewpoints about the hinge structure and motion as deduced from high resolution crystal structures of ovotransferrin N-lobe. *J. Mol. Biol.* 309:937–947.

Navati, M., U. Samuni, P. Aisen, and J. Friedman. 2003. *Bioinorganic Chemistry Special Feature.* Binding and release of iron by gel-encapsulated human transferrin: evidence for a conformational search. *Proc. Natl. Acad. Sci. USA.* 100:3832–3837.

Nolde, S., A. Arseniev, V. Orekhov, and M. Billeter. 2002. Essential domain motions in barnase revealed by MD simulations. *Proteins.* 46:250–258.

Nurizzo, D., H. Baker, Q. He, R. MacGillivray, A. Mason, R. Woodworth, and E. Baker. 2001. Crystal structures and iron release properties of mutants (K206A and K296A) that abolish the dilysine interaction in the N-lobe of human transferrin. *Biochemistry.* 40:1616–1623.

Pakdaman, R., and J. El Hage Chahine. 1996. A mechanism for iron uptake by transferrin. *Eur. J. Biochem.* 236:922–931.

Peterson, N., B. Anderson, G. Jameson, J. Tweedie, and E. Baker. 2000. Crystal structure and iron-binding properties of the R210K mutant of the N-lobe of human lactoferrin: implications for iron release from transferrins. *Biochemistry.* 39:6625–6633.

Peterson, N., V. Arcus, B. Anderson, J. Tweedie, G. Jameson, and E. Baker. 2002. ''Dilysine trigger'' in transferrins probed by mutagenesis of lactoferrin: crystal structures of the R210G, R210E, and R210L mutants of human lactoferrin. *Biochemistry.* 41:14167–14175.

Sanner, M., A. Olson, and J. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 38:305–320.

Schafer, H., A. Mark, and W. van Gunsteren. 2000. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* 113: 7809–7817.

Schafer, H., L. Smith, A. Mark, and W. van Gunsteren. 2002. Entropy calculations on the molten globule state of a protein: side-chain entropies of $\alpha$-lactalbumin. *Proteins.* 46:215–224.

Schlitter, J. 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215: 617–621.

Sharff, A., L. Rodseth, J. Spurlino, and F. Quiocho. 1992. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry.* 31:10657–10663.

Sharma, A., K. Rajashankar, M. Yadav, and T. Singh. 1999. Structure of mare apolactoferrin: the N and C lobes are in closed form. *Acta Crystallogr.* D55:1152–1157.

Sitkoff, D., K. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Srinivasan, J., T. Cheatham III, P. Cieplak, P. Kollman, and D. Case. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.

Steinlein, L., C. Ligman, S. Kessler, and R. Ikeda. 1998. Iron release is reduced by mutations of lysines 206 and 296 in recombinant N-terminal half-transferrin. *Biochemistry.* 37:13696–13703.

Sun, H., H. Li, and P. Sadler. 1999. Transferrin as a metal ion mediator. *Chem. Rev.* 99:2817–2842.

Sutcliffe, M., and N. Scrutton. 2002. A new conceptual framework for enzyme catalysis. Hydrogen tunnelling coupled to enzyme dynamics in flavoprotein and quinoprotein enzymes. *Eur. J. Biochem.* 269:3096–3102.

Taylor, D. 1993. Transferrin complexes with non-physiological and toxic metals. *In* Perspectives in Bioinorganic Chemistry, Vol. 2. JAI Press, Greenwich, CN. 139–159.

Tidor, B., and M. Karplus. 1994. The contribution of vibrational entropy to molecular association. The dimerization of insulin. *J. Mol. Biol.* 238:405–414.

van Aalten, D., A. Amadei, A. Linssen, V. Eijsink, G. Vriend, and H. Berendsen. 1995. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins.* 22:45–54.

Wang, J., P. Morin, W. Wang, and P. Kollman. 2001. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* 123:5221–5230.

Williams, J. 1982. The evolution of transferrin. *Trends Biochem. Sci.* 7:394–397.

Yajima, H., T. Sakajiri, T. Kikuchi, M. Morita, and T. Ishii. 2000. Molecular modeling of human serum transferrin for rationalizing the changes in its physicochemical properties induced by iron binding. Implication of the mechanism of binding to its receptor. *J. Protein Chem.* 19:215–223.

Yang, A., R. MacGillivray, J. Chen, Y. Luo, Y. Wang, G. Brayer, A. Mason, R. Woodworth, and M. Murphy. 2000. Crystal structures of two mutants (K206Q, H207E) of the N-lobe of human transferrin with increased affinity for iron. *Protein Sci.* 9:49–52.

Zak, O., K. Ikuta, and P. Aisen. 2002. The synergistic anion-binding sites of human transferrin: chemical and physiological effects of site-directed mutagenesis. *Biochemistry.* 41:7416–7423.