Thermodynamics of Protein–Peptide Interactions in the Ribonuclease-S System Studied by Molecular Dynamics and Free Energy Calculations†

Thomas Simonson‡ and Axel T. Brünger*

Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Received March 31, 1992; Revised Manuscript Received June 26, 1992

ABSTRACT: Hydrophobic interactions between the S-peptide and S-protein in the ribonuclease-S complex are probed using molecular dynamics simulations and free energy calculations. Three successive mutations at the buried position Met13 are simulated: Met → Leu, Leu → Ile, and Ile → Val, for which X-ray structures and experimental thermodynamic data are available. The calculations give theoretical estimates of the changes in binding free energies associated with these mutations. The calculated free energy differences are small (0–1.6 kcal/mol), in agreement with experiment. However the simulated structures deviate significantly from the experimental ones (mean deviation ~1.5–2 Å), and a large uncertainty in the calculated free energies (1–2 kcal/mol) arises from the multiple minimum problem. Indeed, multiple conformations are available to the side chains around the mutation site, and the sampling of dihedral rotamer transitions is limited, despite long simulations. Fluctuations within each local minimum give rise to a small statistical error. However the uncertainty due to multiple conformations is much greater than the uncertainty due to random statistical errors. In our work, an artificial cancellation of errors arose because we studied conformations of the RNase complex and of the S-peptide that were very similar. In general, the criterion for a precise simulation is not merely to reduce the random statistical error, as has been suggested, but rather to sample all the important local minima along the mutation pathway, and to reduce the statistical error for each one. Our calculations suggest that the packing changes associated with the mutations are energetically small and localized, and largely cancel when the complex and the S-peptide are compared. Solvation of the methionine side chain partial charges in the S-peptide and the complex appear to be energetically equivalent, so that removing them (as in Met13 → Leu, Ile, Val) does not affect binding. Enthalpy and entropy changes could not be estimated reliably.

Internal packing effects are thought to be important in determining protein structure and stability (Richards, 1977). Packing effects also participate in recognition and binding, and their analysis is a step toward rational methods of drug design. Several authors have studied the effect that substitutions of buried hydrophobic residues have on protein stability [Shortle et al., 1990; Kellis et al., 1989; Yutani et al., 1987; Lim & Sauer, 1991; Sandberg & Terwilliger, 1989; Connelly et al., 1990; Varadarajan et al., 1992; Matthews et al., 1992; for reviews, see Alber (1989), Creighton (1991) and Matthews (1991)]. Mutation effects can be very complex, with small changes in stability resulting from large enthalpy–entropy compensation (Shortle et al., 1990) and from many large, mutually cancelling, atomic contributions (Gao et al., 1989), involving both the protein and the solvent, the folded and the unfolded state. Theoretical work can help to understand the details of these microscopic effects. Free energy calculations are one such approach. Thermodynamic perturbation theory (Kirkwood, 1935; Zwanzig, 1954) has been used in conjunction with molecular dynamics to study protein–ligand interactions (Wong & McCammon, 1986; Warshel et al., 1986: Bash et al., 1987) as well as protein stability (Dang et al., 1989; Gao et al., 1989; Tidor & Karplus, 1991; Prévost et al., 1991). For reviews, see Beveridge and di Capua (1989), van Gunsteren and Weiner (1989), Warshel and Aqvist (1989), Jorgensen (1989), and McCammon (1991).

We present here free energy simulations of S-peptide–S-protein binding in the ribonuclease-S system. Bovine pancreatic ribonuclease-A (RNase-A) can be cleaved between residues 20 and 21, giving the S-peptide and the S-protein. These two can be reconstituted to form the complex ribonuclease-S (RNase-S) (Richards, 1955; Richards & Wyckoff, 1970), whose structure (Wyckoff et al., 1967, 1970) is very similar to RNase-A. Only the first 15 residues of the S-peptide contribute to structure and binding (Richards & Wyckoff, 1970; Blackburn & Moore, 1982). Residues 16–20 are disordered in the RNase-S complex, and a 15-residue S-peptide forms a complex almost identical to the 20-residue peptide (Taylor et al., 1981). The hydrophobic residues Phe8 and Met13 of the S-peptide are important for binding (Richards & Wyckoff, 1970). Met13 is almost fully buried in the complex. Connelly et al. (1990) and Varadarajan et al. (1992) have prepared a series of hydrophobic mutations at this position in the 15-residue S-peptide. The free energies, enthalpies, and entropies of binding of the various peptides were measured using titration calorimetry. This approach reveals the effect of the mutations on RNase-S stability without the need to unfold the protein with heat or denaturants. Folding/unfolding is achieved by binding/release of the peptide from the S-protein.

The crystal structures of the native and six mutant complexes—including those simulated here—are known to high resolution (Kim et al., 1992; Varadarajan & Richards, 1992). They are all very similar, with small changes near the mutation site. The rms deviations in backbone positions

† This work was supported by the Pittsburgh Supercomputing Center Grant DMB900079P.

* To whom correspondence should be addressed.

‡ Present address: IBMC, Université Louis Pasteur, 15 rue René Descartes, 67084 Strasbourg, France.

0006-2960/92/0431-8661$03.00/0

Table I: Thermodynamic Parameters for S-Peptide–S-Protein Binding[a]

peptide	calculated $\Delta\Delta G$	observed $\Delta\Delta G$	observed $\Delta\Delta H$	observed $-T\Delta\Delta S$
M → L	1.6	0.1 (0.2)	5.8 (0.8)	−5.7 (0.9)
L → I	−0.8	−0.4 (0.2)	0.0 (0.8)	−0.4 (0.9)
I → V	−0.2	−0.1 (0.2)	−1.0 (0.8)	0.9 (0.9)
M → A		4.0 (0.2)	5.0 (0.8)	−1.0 (0.9)
M → G		5.0 (0.2)	−1.3	6.3

[a] Results are in kcal/mol. T = 298 K. Error bars for the M13G mutant are not known precisely but are greater than for the other mutants.

between the mutant and native structures range from 0.1 to 0.2 Å. The structures of the unbound, 15-residue peptides were monitored by circular dichroism. The binding changes are small (Table I), except for the M13G and M13A mutants. The glycine and alanine side chains are much smaller than Met13, and there appears to be a corresponding cavity in the structures, partially occupied by a water molecule.

We have performed free energy simulations of the mutation of Met13 to three of the residues studied by Varadarajan et al.: leucine, isoleucine, and valine. A 110-ps molecular dynamics simulation of the native protein was also performed. These simulations give detailed predictions about the energy terms and atomic contacts most important for binding. The role of the bound and unbound states are resolved, and the solvent contribution can be separated out. Contributions of individual atomic groups to the free energy changes can also be estimated, as some previous authors have done (Gao et al., 1989; Tidor & Karplus, 1991; Prévost et al., 1991). It should be emphasized, however, that these atomic contributions are not state functions. They depend on the particular computational pathway taken between the initial and final states (see Materials and Methods). Their significance is only qualitative, since different simulation protocols will give different atomic contributions.

Some conclusions about the free energy methodology can be drawn. In particular, we analyze the magnitude and sources of sampling errors. The most important of these is the presence of multiple dihedral conformations. Different simulations usually sample different conformations, leading to "structural hysteresis". Random statistical errors are also discussed.

MATERIALS AND METHODS

Our simulations involve the interconversion of residue 13 from methionine to leucine, to isoleucine, to valine, and each of the inverse steps. To simulate the mutation of a molecule A into a molecule B, the Hamiltonian is progressively mapped from $\mathcal{H}(A)$ to $\mathcal{H}(B)$, along some chosen pathway. We set $\mathcal{H} = \mathcal{H}(\lambda)$, where λ is a coupling parameter that varies from 0 to 1, with $\mathcal{H}(\lambda = 0) = \mathcal{H}(A)$ and $\mathcal{H}(\lambda = 1) = \mathcal{H}(B)$ (Kirkwood, 1935). In the thermodynamic integration approach, the free energy change is expressed as

$$F(B) - F(A) = \int_0^1 \frac{\partial F}{\partial \lambda}(\lambda')\, d\lambda' = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda}(\lambda') \right\rangle_{\lambda'} d\lambda' \quad (1)$$

The subscript λ' indicates an average over the Hamiltonian $\mathcal{H}(\lambda')$. In practice, simulations are done at a few discrete points λ_i along the pathway, and the integral is calculated by interpolation. The choice of pathway, though mathematically immaterial, is practically very important. Thermodynamic integration is formally equivalent to thermodynamic perturbation (Zwanzig, 1954), embodied in the formula

$$F(B) - F(A) = -kT \ln \langle \exp[-(\mathcal{H}_B - \mathcal{H}_A)/kT] \rangle_A \quad (2)$$

but presents practical advantages discussed below.

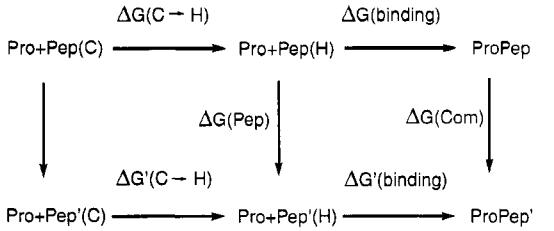

FIGURE 1: Thermodynamic cycle describing S-peptide–S-protein binding. The binding is represented as a two-step process, with the peptide going from its unfolded conformation in solution, Pep(C), to the helical conformation it has in the complex, Pep(H), and then binding to the protein.

To calculate the $\Delta\Delta G$ of binding of a mutant S-peptide to the S-protein, compared to the native peptide, we use the thermodynamic cycle depicted in Figure 1 (Tembe & McCammon, 1984). Simulation of the peptide in solution yields ΔG(Pep); simulation of the complex yields ΔG(Com); whence we obtain

$$\Delta\Delta G = \Delta G'(\text{binding}) - \Delta G(\text{binding}) = \Delta G(\text{Com}) - \Delta G(\text{Pep}) \quad (3)$$

We therefore require a model for the native and mutant S-peptides in solution. We have adopted an α-helical model for the peptide. Specifically, we took the conformation of the peptide as it appeared in the complex. Optical spectroscopy has shown (Connelly et al., 1990) that the native and mutant S-peptides all have about the same helical content in solution, around 10%. We can separate the binding process conceptually into two steps (Figure 1). First the peptide folds into the conformation it will have in the complex (coil-to-helix step), and then it binds to the S-protein. Since the native and mutant peptides all spend about 10% of their time in a helical conformation, the coil-to-helix free energy change is almost the same for all of them: $\Delta G(C \rightarrow H) \approx \Delta G'(C \rightarrow H)$. Therefore, the error introduced by our assumption of a helical peptide will approximately cancel out in the thermodynamic cycle shown in Figure 1. In this work as in the experimental work, the 15-residue S-peptide is used.

Histidine Protonation. The calorimetric measurements (Connelly et al., 1990; Varadarajan et al., 1992) were done at pH 5 and 6 in a phosphate buffer. Results at either pH are similar. Our simulation conditions will mimic pH 6. Histidine 12 titrates with a pK_a of 5.7–5.8, while histidines 119, 48, and 105 titrate at 6.2, 6.3–6.6, and 6.7, respectively (Rüterjans & Witzel, 1969; Meadows et al., 1969; Markley, 1975; Lenstra et al., 1979; Cohen et al., 1980). The histidine protonation state is also sensitive to ligand binding. Binding of substrate analogues stabilizes the doubly protonated forms of His12 and -119, increasing their pK_a's by 0.5–2 units (Meadows et al., 1969; Griffin et al., 1973; Haar et al., 1974). A mixture of singly and doubly protonated forms is found between pH 5 and 8 when a transition-state analogue is bound (Borah et al., 1985). The crystal structures of RNase-S and the mutant complexes were solved at pH 4.75 in an ammonium sulfate solution. In the crystal structure of RNase-S, His12 binds a sulfate ion in conjunction with His112 (Kim et al., 1992). This ion is normally present in the RNase-A crystal structure as well (Wlodawer et al., 1982). Presumably the ion affects the His12 pK_a.

Conversely, the histidine protonation affects the local structure. Structural changes have been detected between pH 5.5 and 6.5 involving this part of the structure, particularly Asp14, Tyr25, His48, and His12 (Blackburn & Moore, 1982). His119 can occupy at least four positions according to the

presence and the nature of ligands at the active site (Richards & Wyckoff, 1970), and it may be sensitive to the state of protonation of His12. The RNase-A structure has been solved in conditions where sulfate was removed from the crystals (Campbell & Petsko, 1987). The differences between the structures with and without phosphate are small. In the latter the bound sulfate ion is replaced by a water molecule.

At pH 6, His12 probably occupies a mixture of states, including a doubly protonated state with a bound phosphate ion and a neutral state with a bound water molecule. As our model, we took a neutral His12, in the $N^{\delta 1}$–H tautomer (Walters & Allerhand, 1980), with no bound ion. The other histidines were assumed doubly protonated. This model should be reasonable for pH 6.

Energy Parameters. The OPLS/AMBER force-field was used to model the protein (Jorgensen & Tirado-Rives, 1988; Weiner et al., 1984), and the TIP3P model was used for the solvent (Jorgensen et al., 1983). Polar hydrogens are explicitly included in this model, and hydrogen bonding is modeled as a purely electrostatic interaction. Induced dipoles are not included in the model. This is reasonable, as the mutations do not involve major changes in the electrostatic charge distribution. Electrostatic interactions were truncated on a group basis at a distance of 9 Å. Groups were made up of the amino acid side chain portions, the backbone portions, and individual water molecules. We used a dielectric constant of 1.

Stochastic Boundary Molecular Dynamics. Molecular dynamics simulations were carried out with the stochastic boundary method (Brooks & Karplus, 1984; Brooks et al., 1985; Brünger et al., 1985). They included an approximately spherical region centered on the β-carbon of the residue being mutated, with a radius of 14 Å. In the protein simulations, this region contained about half of the protein and 172–207 water molecules, depending on the simulation. No residues are subdivided, i.e., all residues in the simulation region are intact, although the peptide chain is broken in places. No counterions were included in the simulations. In all cases the simulated region was electrically neutral. For each native or mutant protein considered, the initial model was the appropriate X-ray structure (Kim et al., 1992; Varadarajan & Richards, 1992). The C-terminus of the S-peptide is amidated. Protein atoms in the outer 2 Å of the simulation region were restrained by a harmonic force to stay close to their initial positions. The exact restraining positions varied with the coupling parameter λ, being a linear interpolation between the X-ray coordinates of the starting molecule ($\lambda = 0$) and those of the final molecule ($\lambda = 1$). For each S-peptide, starting coordinates were extracted from the X-ray structure of the corresponding protein–peptide complex. The 14-Å simulation region then encompassed the whole peptide. The backbone of Lys1 was restrained harmonically in the way just described, to avoid drifting of the peptide. In addition, after several trial simulations, weak harmonic pseudobonds were introduced in the middle of the peptide to maintain the helical hydrogen bonds. These restraints will not affect the calculated free energy differences significantly.

The second-order, nonbonded list-based, Verlet algorithm was used to integrate the equations of motion (Verlet, 1967). Temperature coupling (Berendsen et al., 1984) was used to keep the temperature at 298 K, the temperature of the experiments. The pressure remains approximately constant since the mutations we studied have small volume and chemical changes. The geometry of each water molecule was kept rigid with the SHAKE algorithm (Ryckaert et al., 1977), and hydrogen masses were set to 10 amu so that a time-step of 2 fs could be used to integrate the equations of motion (Pomes & McCammon, 1990). Artificial masses affect the dynamic properties but not the configuration integral of the system. The latter completely determines the free energy change $\Delta\Delta G$; any contributions of the atomic momenta cancel out in the two sides of the thermodynamic cycle in Figure 1.

In addition to the free energy simulations, a 110-ps simulation of the native structure was done. All simulations were run with the program X-PLOR (Brünger, 1987), on Convex C220 and Cray YMP computers.

Choice of ΔG Method. Almost all of our calculations use the thermodynamic integration method; one portion uses the thermodynamic perturbation method (Beveridge & di Capua, 1989). Thermodynamic integration permits a better control over the data collection and permits a decomposition of the free energy into atomic contributions (which depend, however, on the choice of the simulation pathway). Thermodynamic integration and perturbation differ by terms on the order of $\delta\lambda^3$ at most, where $\delta\lambda$ is the increment of the coupling parameter between windows (T. Simonson, unpublished results). Our coupling parameter has constant increments of about 0.1–0.2, so that $\delta\lambda^3$ is quite small.

At the start of each mutation ($\lambda = 0$), random velocities were drawn from a Maxwellian distribution at 298 K, and 40 ps of equilibration was performed. At subsequent values of λ, ~6 ps of equilibration was performed, followed by 10–20 ps of data collection. Ordinarily, 11 steps for λ between 0 and 1 were considered. Thus each mutation involved 170–280 ps of molecular dynamic simulation. Thermodynamic integration allows one to add intermediate data points along the mutation pathway a posteriori, if the statistical error is too large for certain values of λ or if the free energy derivative varies so rapidly that accurate interpolation is impossible. This never occurred in our mutations. Each mutation was performed forward and backward. The initial structure in each direction was always the relevant X-ray structure.

We simulated the interconversion of residue 13 from methionine to leucine, to isoleucine, and to valine and each of the inverse steps. Each mutation involves the creation or deletion of one or two atoms. The Met ↔ Leu mutations involve adding or removing partial charges on the methionine C_γ–S_δ–C_ϵ. In each mutation, the minimum number of atoms is used; for example, in the Met → Leu case, Met-SD is mapped into Leu-CD1, while Met-CE is removed and Leu-CD2 is added.

When a new atom is introduced into a structure, a very rapid variation of the free energy can occur unless one proceeds very gradually. This is the well-known "origin catastrophe" (Beveridge & di Capua, 1989). To avoid this problem, appearing atoms were introduced with the bond to their nearest neighbor initially shortened to ~0.5 Å (Pearlman & Kollman, 1989). The appearing atom is thus shielded by its neighbor, so that other atoms cannot pass through it freely. The bond length is increased to its true length progressively, either while the van der Waals weight of the atom is increased, or after it has been increased to its full value.

The van der Waals well-depth ϵ and the changing bond length were varied in a linear way, usually in 10 steps. Thus, the van der Waals contribution to the free energy derivative is

$$U_{vdW}(\lambda) = \lambda U_{vdW}(1)$$

$$\left\langle \frac{\partial U_{vdW}}{\partial \lambda}(\lambda) \right\rangle_\lambda = \langle U_{vdW}(1) \rangle_\lambda \qquad (4)$$

The free energy derivative is given by the full van der Waals energy of the appearing atom, averaged over a simulation with weight λ. The bond contribution is

$$\left\langle \frac{\partial U_{\text{bond}}}{\partial \lambda}(\lambda) \right\rangle_{\lambda} = -2k_{\text{bond}}[\langle b \rangle_{\lambda} - b_0(\lambda)][b_0(1) - b_0(0)] \quad (5)$$

The free energy derivative is given by the mean stretching of the bond. An alternative approach would be to calculate the mean force along the bond due to the surrounding atoms and derive the work associated with incrementing its length (Tobias & Brooks, 1987; Anderson & Hermans, 1988). Changes in atom types, and slight changes in angular or dihedral parameters, also enter into the mutations and are treated in a straightforward way. Integration of the free energy derivatives was done here with a simple trapezoidal approximation. Comparison to a cubic spline approximation was done in several cases to estimate integration errors. The detailed mutation pathways are described in the Appendix.

Random Statistical Errors. Among the different sources of error, one of the easiest to estimate is the random statistical error (Davenport, 1970; Schiferl & Wallace, 1985; Straatsma et al., 1986). A time series $\{a_i; i = 1, n\}$ of the quantity of interest (e.g., U_{vdW}) is taken from the molecular dynamics trajectory. The standard error $\sigma(\bar{a})$ of the time average $\bar{a}$ depends not only on the size n of the sample but also on the correlations within the sample (Davenport, 1970):

$$\sigma(\bar{a})^2 = \sigma(a)^2/n + \frac{2}{n}\sum_{l=1}^{n-1}(1 - l/n)K(l) \quad (6)$$

where $\sigma(a)$ is the standard deviation of the time series $\{a_i\}$ and $K(l)$ is the time-correlation function

$$K(l) = \overline{a_i a_{i+l}} \quad (7)$$

$K(l)$ can be estimated from the trajectory.

Dial Plots. A more severe source of error in the case of protein calculations is structural hysteresis. Several packing arrangements are available to the structure, corresponding, for example, to different sets of dihedral rotamers (Janin et al., 1978; Ponder & Richards, 1987). Repeated simulations of a given mutation do not necessarily sample the same set of rotamers. To monitor the spontaneous jumps between rotamers, we have produced "dials" plots, similar to those of the program Dials and Windows (Swaminathan et al., 1990). For each dihedral angle of interest, the instantaneous value $\phi(t)$ is represented by a point in a small circular dial. The radial direction measures time t, and the polar angle measures $\phi(t)$ (Figure 7). Our plotting routine is interfaced with X-PLOR.

Decomposition of ΔΔG. The ΔΔG of binding can be separated conceptually into two parts (Prévost et al., 1991). One part arises from the changes in solvation of the system upon mutation. Another part is due to changes occurring in vacuo. Indeed,

$$\Delta\Delta G^{(\text{wat})} = \Delta\Delta G^{(\text{vac})} + [G^{\text{solv}}(\text{Pep}') - G^{\text{solv}}(\text{Pep})] - [G^{\text{solv}}(\text{ProPep}') - G^{\text{solv}}(\text{ProPep})] \quad (8)$$

$G^{\text{solv}}(\text{A})$ is the transfer free energy of A from vacuum to water. The first term on the right-hand side is the ΔΔG in vacuo. The term in brackets is the change in solvation free energy of the peptide upon mutation minus the change in solvation free energy of the complex upon mutation. We have performed the Met13 → Leu mutation in vacuo for the purpose of this analysis.

An advantage of thermodynamic integration is that the free energy change can be decomposed into a sum of atomic contributions. However, these contributions are not state functions, and the decomposition will depend on the pathway taken between the initial and final states.[1] Therefore, the decomposition has mainly a qualitative value, suggesting typical energy magnitudes and a typical interaction range, but does not represent a quantitative prediction that can be related to the effects of second-site mutations, say.

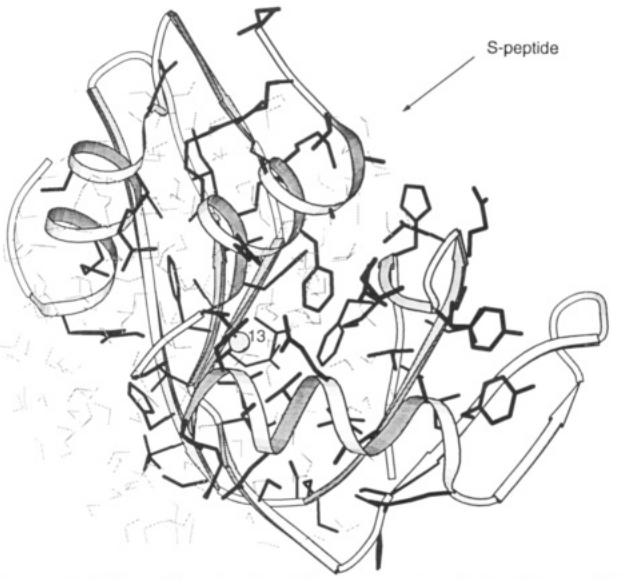

FIGURE 2: RNase-S molecule: ribbon trace of entire backbone; side chains in the simulated region (black); simulated water molecules (grey). The Met13-SD atom is shown as a sphere and labeled "13". This plot was produced with the program Molscript (Kraulis, 1991).

Separation of the free energy changes into enthalpy and entropy components can be done in theory, using the time average of the potential energy to approximate the enthalpy. In practice we found this approach to be useless, since the enthalpy changes are obtained as small differences between large numbers that have a large statistical uncertainty. The finite-differences expression of Brooks (1988) for the enthalpy may give better results, but we did not test it. Nor do our simulations give information about heat capacity changes or changes in the protein dynamics associated with the mutations, because of the artificial hydrogen masses and the simulation of a subset of the protein.

RESULTS

Simulation of the Native Structure. The C_α trace of the native X-ray structure is shown as a ribbon plot in Figure 2. The side chains included in the simulation are drawn in bold. Simulated water molecules are shown in thin lines. After equilibration and 110 ps of molecular dynamics, the rms deviation from the X-ray structure is 1.2 Å for the heavy atoms, 0.75 Å for the backbone, and 1.36 Å for the side chains. These differences arise from (1) inaccuracy of the potential function, (2) errors in the refined atomic positions, and (3) lack of crystal contacts in the simulation. The secondary structure and the overall shape and fold of the solution structure are known to be close to those of the crystal structure (Rico et al., 1989, 1991; Robertson et al., 1989). However, in the

[1] It is obvious that the atomic contributions are not state functions in general. Consider, for example, a finite one-dimensional chain of atoms with nearest neighbor interactions; consider two different pathways for changing the chemical type of one atom of the chain: one in situ and one where the atom is first moved to the end of the chain, then changed, then moved back. The second pathway gives a nonzero contribution for the atom at the end of the chain; the first pathway does not.

crystal, 43% of the molecular surface is involved in crystal contacts (the solvent content is about 40%), so that differences at the atomic level are to be expected. The estimated error in atomic positions in the crystal structure is about 0.05 Å for backbone atoms and 0.34 Å for side chains (viz., one standard deviation of the positions from a series of repeated refinements; Kim et al., 1992). Since the uncertainty of atomic positions is low in the protein core, the uncertainty on surface atoms is greater than this average error bar. Figure 3a shows the residue-by-residue deviations. The largest deviations involve residues 14, 15, and 48. Residues 14 and 15 are very mobile in the simulation. Ser15 is partly disordered in the X-ray structure. In the RNase-S complex with the 20-residue S-peptide, residues 16–20 are completely disordered (Kim et al., 1992). Therefore, this part of the chain can move a good deal at little energetic cost. In the crystal structure, Asp14 is involved in a crystal contact. Asp14-OD2 is hydrogen bonded to Tyr25-OH, while Asp14-OD1 makes a hydrogen bond to Lys91-NZ in a symmetry-related molecule. The Asp14–Tyr25 hydrogen bond exists in all the refined RNase-A and RNase-S crystal structures (Kim et al., 1992), which cover a broad range of pH's. In the simulation, the lattice partner Lys91 is absent, and the hydrogen bond to Tyr25 is broken. The structure rearranges itself during the equilibration stage to form the hydrogen bonds Asp14-OD2–His48-ND1 and Asp14-OD1–Arg33-NH1. In the M13L, M13I, and M13V crystal structures, the Asp14-OD2–His48-ND1 distance is only 3.3–3.4 Å (Varadarajan et al., 1992b). In our simulations of the mutants, His48 approaches Asp14 during equilibration and forms an Asp14-OD2–His48-ND1 hydrogen bond. In the simulation of the M13L structure, Asp14-OD2 hydrogen bonds to both Tyr25-OH and His48-ND1. In the simulation of the M13I structure, Tyr25-OH shifts away 1.5 Å, and its hydrogen bond to Asp14 is broken. In the simulation of the M13V structure, Tyr25-OH shifts away 3.5 Å. The approach of His48 is not surprising. Asp14 and His48 bear full and opposite atomic charges; their side chains are about 70% buried (solvent-inaccessible) in the X-ray structures, so that the effective local dielectric constant is probably moderate, and there is a large free energy gain in their approach, probably on the order of 1 kcal/mol. The atomic charges are omitted in the X-ray refinements (Kim et al., 1992; Varadarajan et al., 1992b). The M13L, M13I, M13V, and native crystal structures are overall very similar, with mutual rms deviations of 0.1–0.2 Å for backbone atoms. Therefore, it may be possible to view their Asp14–Tyr25–His48 arrangements as alternate structures available to the native complex in solution at little or no energetic cost.

Nuclear magnetic resonance also gives some evidence for a hydrogen bond from Asp14 to His48 and not Tyr25 in solution at acidic pH. The proton resonances of the Tyr25 ring have been assigned, and they suggest that, in acidic solution, Tyr25 does not form a hydrogen bond (Lenstra et al., 1979), although some authors have contested this (Niu et al., 1979), on the basis of ^{13}C NMR studies. ^{13}C NMR studies with an enriched S-peptide show an inflection of the Asp14 titration curve at pH 6.1 in RNase-S, thought to arise from His48 (Niu et al., 1979; Cohen et al., 1980). Several authors have suggested that an Asp14–Tyr25 hydrogen bond exists in solution above pH 6.5 in RNase-S, but that when His48 is protonated, it replaces Tyr25 as a hydrogen-bond donor to Asp14 (Rüterjans & Witzel, 1969; Santoro et al., 1979). The data of Lenstra et al. (1979) indicate large motions of Tyr25 around its χ_2 dihedral, where are seen in the simulation but not in the crystal structures, where Tyr25 is restricted by crystal contacts. Thus, the rearrangements seen in the simulation may reflect elements of a real alternative structure, which may be partially present, or even predominant, in the conditions of the calorimetric experiments.

The simulated rms atomic fluctuation, averaged over the protein heavy atoms, is 0.70 Å. The experimental value, from the crystallographic *B* factors (Kim et al., 1992), is 0.85 Å. This includes overall rotation–translation of the molecule, as well as static disorder. Howlin et al. (1990) estimated the overall rotation and translation amplitudes in RNase-A crystals to be 0.64 and 0.24 Å. Taking this along with the molecular dynamics estimate of intramolecular motions, we get an estimated total rms displacement of 1.08 Å. The calculated variation of the mobility along the polypeptide chain agrees roughly with the distribution of *B* factors (not shown). Several residues that are disordered in the X-ray structure are in or near the boundary zone of the simulation, which is harmonically restrained.

The simulation zone includes 36 crystallographic water sites. Twenty-four of these are occupied during most of the simulation. Six are occupied by the same water molecule throughout. These six sites involve two hydrogen bonds to the protein on average, one being to a charged residue. Two of the crystallographic waters are found in all the mutant crystal structures as well (Varadarajan et al., 1992b), making hydrogen bonds to Asp53, Glu49, and Ser50 and to Asp83, Lys98, and Thr100. In the simulation, Asp53 and Asp83 bind the same water molecule throughout the simulation, although in a different geometry from the crystal structure.

Several side chains are quite mobile in the simulation, including the Asp14, Ser15, Tyr25, and His48 side chains. The large motions of Tyr25 around its χ_2 dihedral agree with the NMR results of Lenstra et al. (1979). His119 occupies the single conformation seen in the crystal structure throughout the simulation. There is no hydrogen bond between His119 and Asp112 in either the crystal structure or the simulation, contrary to some previous suggestions (Blackburn & Moore, 1982) and in agreement with earlier simulations of RNase-A (Brünger et al., 1985). The closest approach of any atom of His119 to His12 is 5.5 Å, contrary to the direct interaction postulated in the Witzel reaction mechanism (Witzel, 1963).

The Met13 side chain is almost completely buried, and consequently its solvent-accessible area is small (between 0 and 5 Å^2) throughout the simulation. Partial charges on the methionine side chain, in this model, are as follows: C_γ, 0.235 au; S_δ, –0.47 au; C_ϵ, 0.235 au. The electrostatic interaction energy of the side chain with the solvent is about –5 kcal/mol on average.

Calculated and Observed Binding Free Energies and Structures. The calculated and observed (Connelly et al., 1990; Varadarajan et al., 1992a) $\Delta\Delta G$'s for the three simulated mutations are given in Table I. The agreement is similar to other recent free energy calculations [e.g., Tidor and Karplus (1990) and Prévost et al. (1991)]. The calculated results are averages over forward and backward simulations of each mutation. The $\Delta\Delta G$'s from repeated simulations of the same mutation differ by 0.2–0.4 kcal/mol. That these differences are so small is probably fortuitous. The uncertainty in the simulated results is actually closer to 1–2 kcal/mol. This is larger than the range of observed affinities. The calculations rank the four peptides incorrectly in order of decreasing affinity for the S-protein: M13M > M13V > M13I > M13L; the experimental ranking is M13V > M13I > M13M > M13L.

The simulated mutant structures deviate significantly from the X-ray structures, as shown in Figures 3 and 4 and

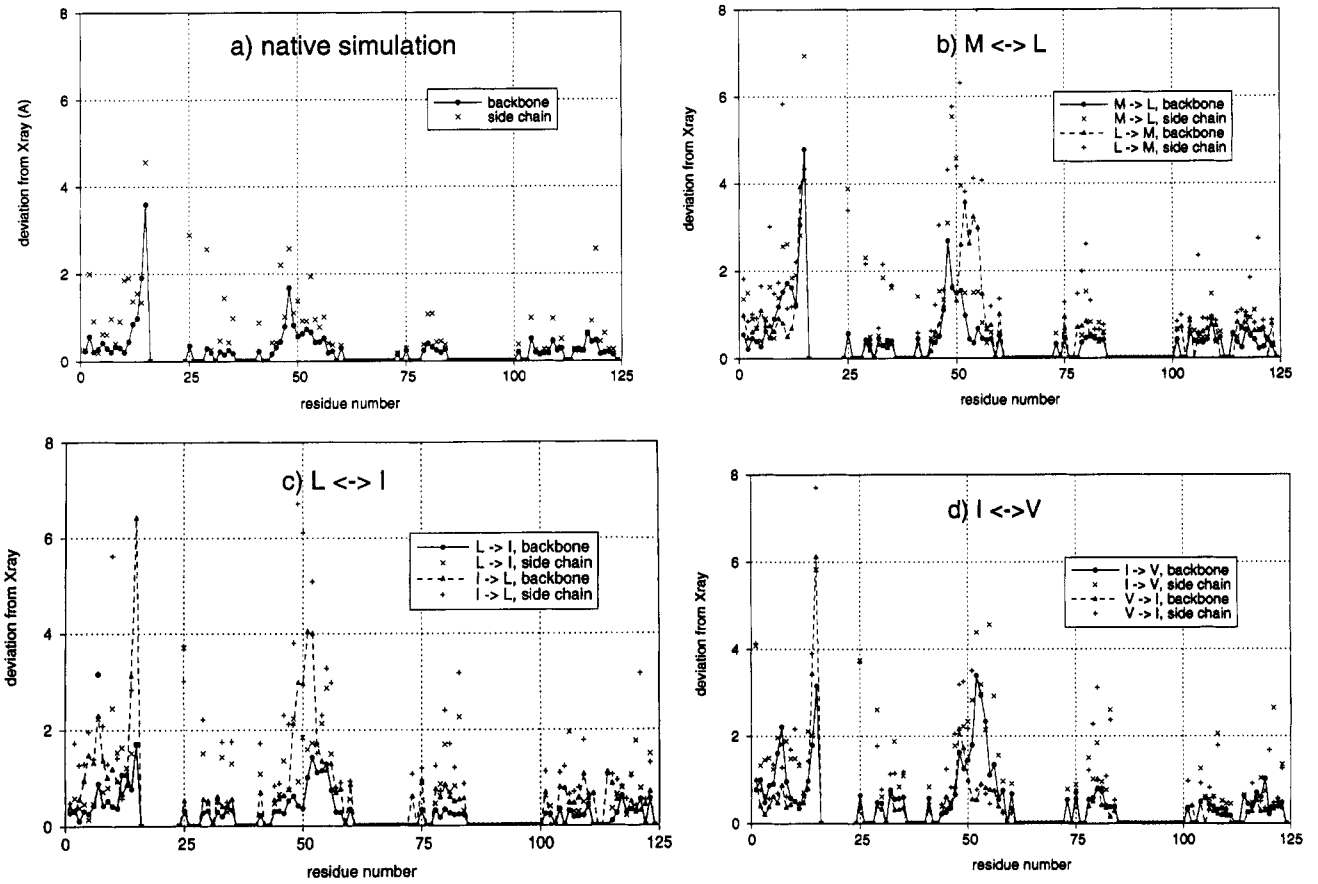

FIGURE 3: rms deviations of the native and mutant simulations from the corresponding crystal structures, averaged over each residue. (a) Native simulation; (b) Met ↔ Leu simulated structures; (c) Leu ↔ Ile simulated structures; (d) Ile ↔ Val simulated structures.

Table II: Deviation of Simulated Mutant Structures from Crystal Structures[a]

mutation	native[b]	M→ L[c]	L→ M	L→ I	I→ L	I→ V	V→ I
heavy atoms	1.2	1.7	2.0	1.2	2.3	1.7	1.7
backbone	0.8	1.2	1.4	0.6	1.7	1.3	1.2
side chains	1.4	1.9	2.3	1.5	2.6	1.9	2.0
inner sphere[d]	1.3	2.0	2.6	1.4	2.8	1.9	2.0
outer sphere	1.0	1.1	1.2	1.1	1.4	1.4	1.3

[a] Results are in angstroms. [b] Native simulation compared to the crystal structure. [c] Structure obtained by simulating the Met13 to Leu mutation, compared to the M13L crystal structure. [d] Inner 6 Å of the simulation region. The rest is the "outer sphere".

summarized in Table II. As in the native simulation, the largest deviations are at the end of the S-peptide and in the region of residues 48–56. Val47 and Leu51 are in direct contact with Met13. In the native structure obtained at the end of the Leu → Met simulation, large deviations occur at positions 52–56, within the mixed $3_{10}/\alpha$-helix 50–60. The end of the helix (57–60) is kept in place by harmonic restraints since it is at the boundary of the simulation region. The stretch between (not including) the helical hydrogen bonds Ser50-O–Val54-H and Ala56-O–Gln60-H has unwound, so that residues 54, 55, and 56 have flipped out of the helix (Figure 4a). The C-terminus of the S-peptide has swung around. In the reverse direction, in the simulated M13L structure, the backbone has rotated around $\phi(47)$ and $\phi(50)$, giving a packing arrangement completely different from the crystal structure (Figure 4b). Deviations of a similar magnitude occur in the M13I and M13V simulated structures.

Sampling Errors and Structural Hysteresis. The calculated $\Delta\Delta G$'s can be separated into their components ΔG(Pep) and ΔG(Com), corresponding to the unphysical processes Pep → Pep′ and ProPep → ProPep′. This is reported in Table III. The random statistical errors are shown in parentheses (1σ). ΔG(Pep) and ΔG(Com) are much less reproducible then the $\Delta\Delta G$'s, with differences of over 2 kcal/mol between different simulations of the same mutation. These differences arise mainly from different conformations sampled by different simulations. They happen to cancel out when $\Delta\Delta G$ is taken. This occurs in part because, in our protocol, the S-peptide always has the same starting structure in the complex and in solution, giving similar local packing interactions in the two sides of our thermodynamic cycle.

ΔG(Pep) and ΔG(Com) can be further subdivided by distinguishing the most important terms in the Hamiltonian. In the Met13 ↔ Leu mutations, the three most important terms are (1) the interaction of the Met13 side chain partial charges with their surroundings, (2) the van der Waals interactions of the CE and CD2 groups with their surroundings, and (3) the bond energy associated with the lengthening/shortening bonds to the CE and CD2 groups. Figure 5 and Table III show the contribution of each of these terms to ΔG(Pep) and ΔG(Com). Note that these contributions are not state functions and depend on the simulation pathway. The contribution to $\Delta\Delta G$ is the small difference between ΔG(Pep) and ΔG(Com). The data in Figure 5 are an average over one forward and one backward simulation. The error bar for each point in the figure is plus or minus two standard errors (random statistical error as described under Materials and Methods). The relevant correlation functions have only roughly converged, so that the statistical error estimates are approximate. For some windows, due to lack of convergence of the correlation function, we obtain a negative variance of the free energy derivative. In this case, we assigned an arbitrary error bar, equal to the largest error bar found in any

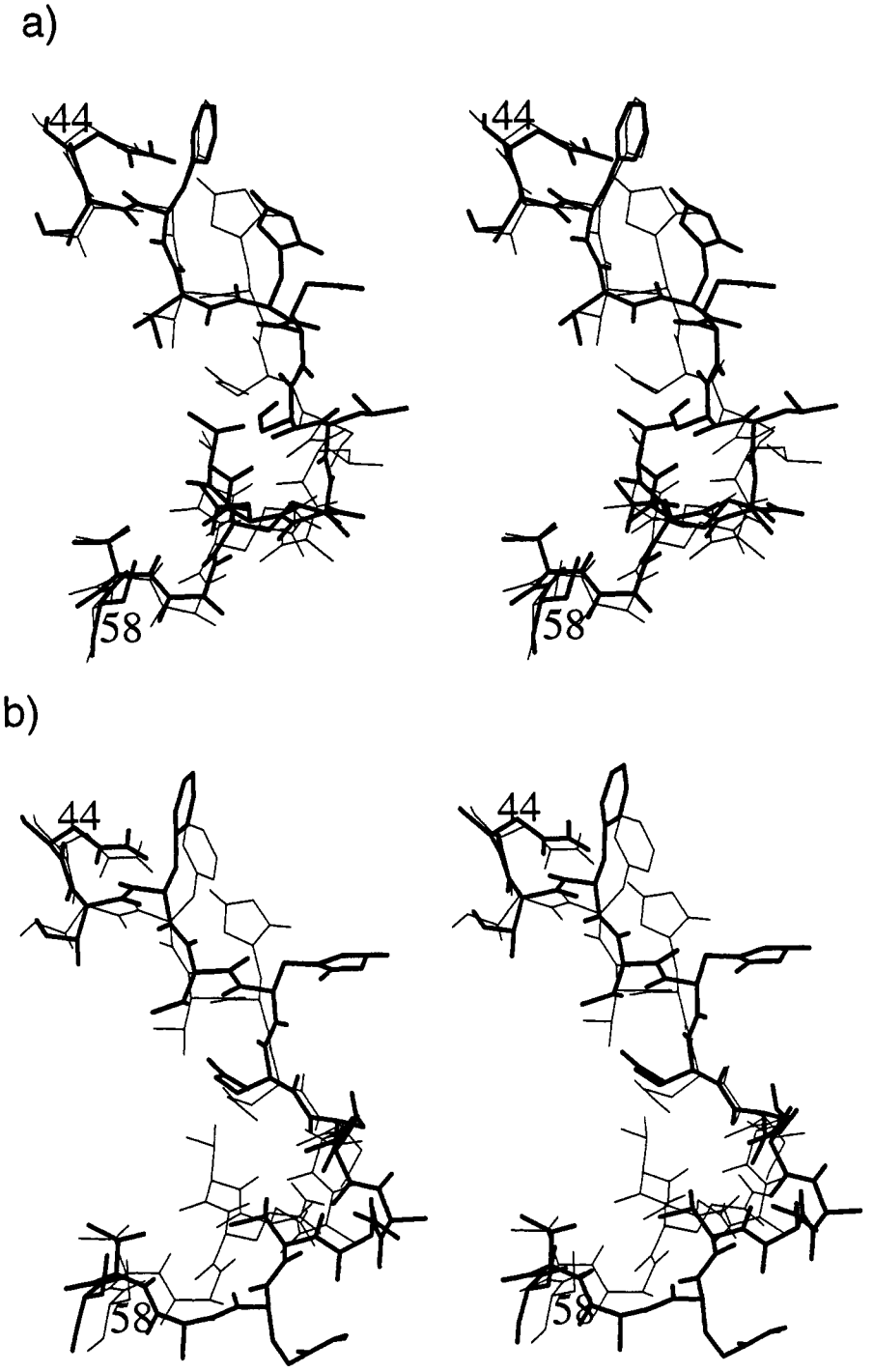

FIGURE 4: Stereoview of portions of the simulated crystal structures: (a) M13M simulated (Leu ↔ Met simulation; bold) and X-ray (thin) structures, 44–58 region. (b) M13L simulated (bold) and crystal (thin) structures, backbone region from residue 44 (top) to 58 (bottom), and Val47 side chain.

of the other windows. The bond terms have relatively large statistical errors, since the sampling of the trajectory was done inadvertently at a frequency (every 0.04 ps) that is close to the natural frequency of the bonds, giving large correlations in the relevant time series. The van der Waals and electrostatic terms, on the other hand, have small random errors.

A forward and backward mutation of the Pro–Pep complex are compared in Figure 6, to illustrate the hysteresis associated with each of the three energy terms. The hysteresis is quite small for the electrostatic term, despite the different structures sampled in the two calculations. For the bond and van der Waals terms, the hysteresis values are about 1 kcal/mol and of opposite signs. This is fairly typical: in several cases a large van der Waals contribution in a forward simulation will be replaced by a large bond contribution in the backward simulation, as the two simulations sample different structures with different local interactions (Table III).

Figure 7 is a dials plot showing some of the structural differences between the Met13 → Leu simulation and the reverse simulation. This plot illustrates some of the problems caused by multiple dihedral conformations and structural hysteresis. In the Met13 → Leu simulation, Met/Leu13 remains in its starting conformer throughout, (χ_1, χ_2) = (gauche$^-$, gauche$^-$). It never jumps into the experimental Leu13 conformer, (trans, gauche$^+$). In the reverse simulation, residue 13 starts out in the (trans, gauche$^+$) conformer. χ_1 jumps immediately to gauche$^-$, and then χ_2 jumps to trans, and eventually χ_1 jumps to gauche$^+$. The final conformation,

Table III: Hysteresis of the Main Energy Terms in ΔG(Pep) and ΔG(Com)[a]

	mutation	M → L	L → M	L → I	I → L	I → V	V → I
ΔG(Pep)	elec	2.0 (0.2)	−2.1 (0.1)				
	vdW	2.8 (0.2)	−3.8 (0.1)	−2.9 (0.1)	−1.2 (0.2)	−2.2 (0.1)	0.9 (0.3)
	cov[b]	0.0 (0.9)	2.7 (–)	1.5 (–)	−0.9 (0.7)	0.2 (0.3)	1.6 (0.9)
	total	4.8 (0.9)	−3.1 (–)	0.5 (–)	−2.1 (0.7)	−2.1 (0.3)	2.6 (0.9)
ΔG(Com)	elec	2.2 (0.2)	−2.4 (0.1)				
	vdW	2.7 (0.2)	−3.0 (0.1)	−0.3 (0.2)	−0.2 (0.4)	−2.4 (0.0)	2.7 (0.1)
	cov	1.5 (1.8)	0.9 (–)	−0.2 (–)	−1.1 (0.7)	0.4 (0.4)	0.1 (0.5)
	total	6.5 (1.8)	−4.6 (–)	−0.2 (–)	−1.2 (0.8)	−2.0 (0.4)	2.9 (0.5)
$\Delta\Delta G$	elec	−0.2 (0.3)	0.3 (0.1)				
	vdW	0.1 (0.3)	−0.8 (0.1)	1.6 (0.2)	1.1 (0.4)	−2.4 (0.1)	2.7 (0.3)
	cov	−1.5 (2.0)	1.8 (–)	−1.7 (–)	−0.2 (1.0)	−0.2 (0.5)	1.5 (1.0)
	total	1.7 (2.0)	−1.5 (–)	−0.6 (–)	0.9 (1.1)	0.1 (0.5)	0.3 (1.0)

[a] Results are in kcal/mol. [b] Covalent energy terms. Standard errors in parentheses.

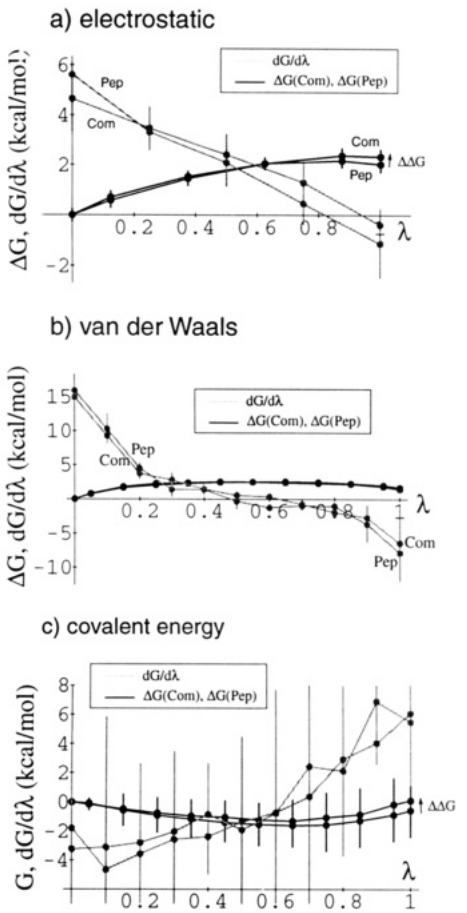

FIGURE 5: Free energy derivatives (grey) and free energy changes ΔG(Com) and ΔG(Pep) (black) vs λ; Met13 → Leu simulation; RNase complex compared to S-peptide. (a) Contribution of Met13 side chain charges; (b) contribution of the CD2 and CE van der Waals interactions; (c) contribution of the bond stretching terms. Results for the complex are labeled "Com"; results for the S peptide are labeled "Pep". The results are averaged over one forward and one backward simulation. Error bars represent twice the random statistical error (see eq 6). In panel b the error bars for ΔG are too small to be visible.

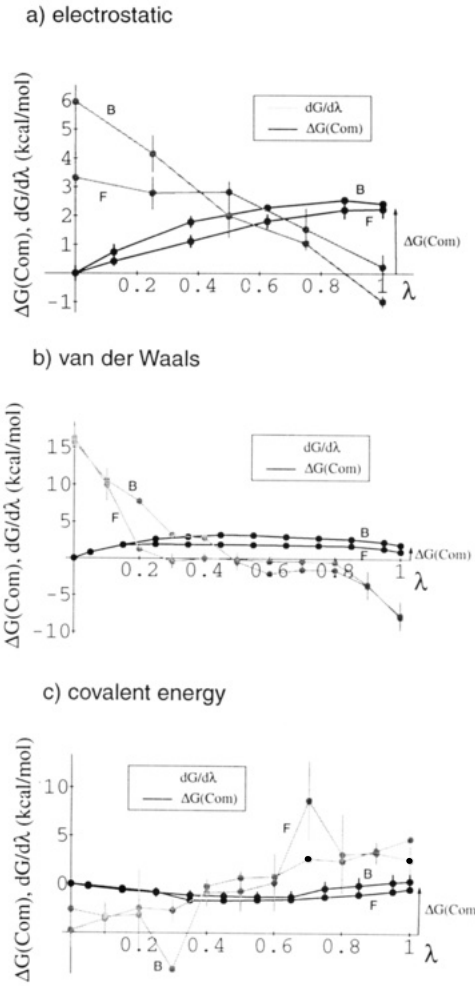

FIGURE 6: Hysteresis in the Met13 ↔ Leu calculations of ΔG(Com); Met13 → Leu calculation compared to reverse calculation; RNase complex; free energy derivative (grey) and integrated free energy change ΔG(Com) (black) vs λ. (a) Contribution of Met13 side chain charges; (b) contribution of the CD2 and CE van der Waals interactions; (c) contribution of the bond stretching terms. Forward results are labeled F, backward results B. In the backward calculation, λ was varied from 0 to 1. The backward results are translated along the vertical axis to make the F and B curves coincide at λ = 0. Error bars represent twice the statistical error (see eq 6). In panel b the error bars for ΔG are too small to be visible. In panel c the "backward" plots do not include error bars.

(gauche$^+$, trans), is distinct from both the native and the mutant experimental conformations. Several other important residues visit different conformers in the forward and reverse simulations: Gln11, His12, His48, Glu49, and Val54, for example. The backbone structure in the region 47–50 is very different in the two simulations, though very similar in the two crystal structures. The local geometry deteriorates somewhat in the course of the Leu → Met simulation, and the Met/Leu13 χ_1 and χ_2 angles fluctuate wildly, suggesting that the structure is trapped in a high-energy conformation and cannot relax into the structure of the native complex.

Contribution of Individual Atoms and Groups to Binding. One of the advantages of thermodynamic integration as opposed to thermodynamic perturbation is that the free energy is expressed as a linear function of the Hamiltonian, so that the contribution of individual atomic groups can be determined

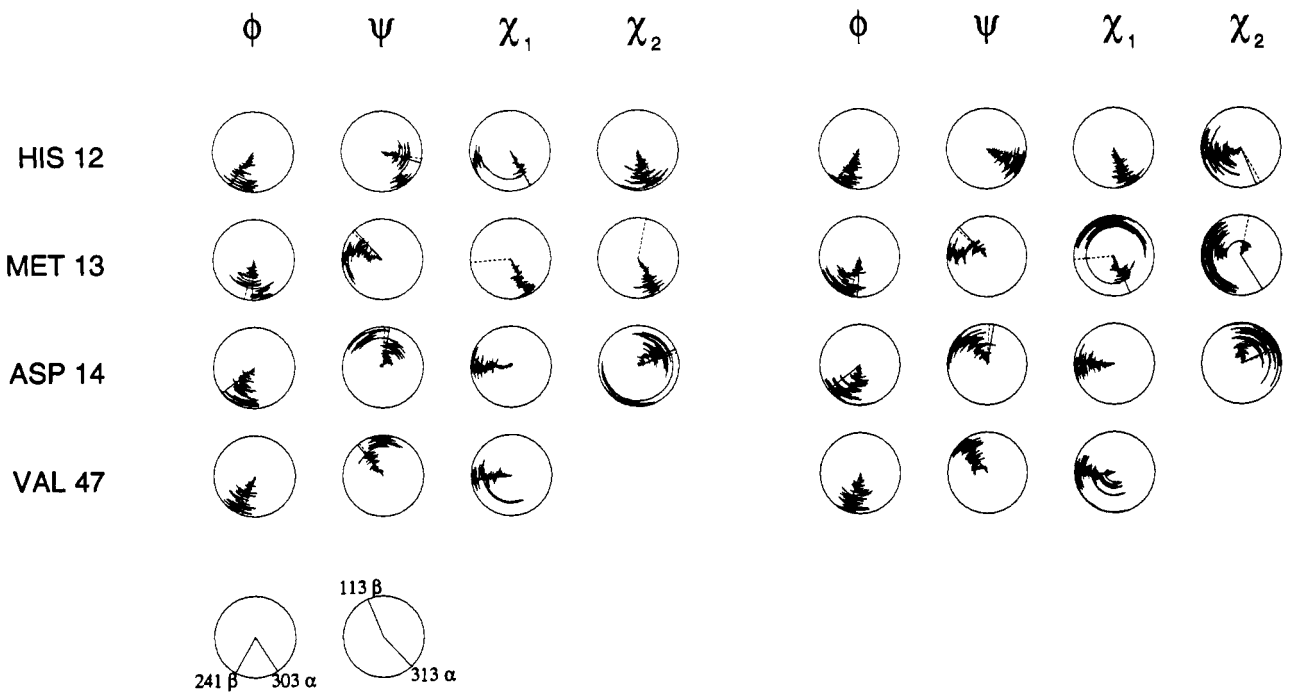

FIGURE 7: Dials plot of the dynamics of several residues in the RNase complex during the Met13 → Leu and the reverse mutations. In each dial, a single dihedral angle is shown as a function of time. The center of the dial corresponds to the beginning of the simulation; the outer edge corresponds to the end of the simulation. The two dials at the bottom of the figure indicate typical values of ϕ and ψ for an α-helix and a β-sheet, respectively. The zero angle is at three o'clock, and the positive sense is counterclockwise. The native X-ray structure is shown in each dial as a straight solid line; the M13L X-ray structure is shown as a dashed line.

(Gao et al., 1989). These contributions—unlike the total free energy—are not state functions. They depend on the pathway chosen between the initial and final states, and a different simulation protocol would give a different decomposition. They merely give typical energy magnitudes and a typical range of interactions. We distinguish (1) intraresidue energy terms, within residue 13, including covalent and nonbonded interactions, (2) interactions with several important protein residues, and (3) interactions with the solvent. The decomposition is summarized in Table IV for all three mutations.

In contrast to mutations involving the creation or redistribution of net charges (Gao et al., 1989; Tidor & Karplus, 1991), the free energies here can be attributed to just a few atomic groups in the immediate vicinity of the mutation. This reflects the short range of the van der Waals interaction. Residues 8–10, 12, 14, 47, and 51 are all in direct contact with position 13 at some time during the simulations, in either the S-peptide or the complex. This decomposition does not mean that more distant parts of the structure do not contribute indirectly to the free energy, through interactions with the residues surrounding position 13. The solvent accounts for much of the cost of removing the Met13 partial charges, while the protein and solvent both contribute to the free energies through the small van der Waals terms.

The free energy for removing the Met13 partial charges is 2 kcal/mol, in both the complex and the peptide in solution. The interaction energy of these partial charges with their surroundings is −5 kcal/mol in the native structure. The difference between the interaction energy and the interaction free energy underlines the essential difference between free energy calculations and energy estimates based on a single, "end-point" structure, as pointed out by other authors (Tidor & Karplus, 1991). Here, the difference between the two energies represents a dielectric relaxation energy. The series of simulations with decreasing partial charges simulate the dielectric relaxation of the structure in response to the charge removal.

Solvent Contribution to Binding. The contribution of the solvent to the $\Delta\Delta G$ of binding can be extracted, as described under Materials and Methods, by simulating the mutation in vacuo as well as in solution (eq 8). This contribution is distinct from the solvent contribution defined in the previous paragraph (Table IV). In Table IV we merely sum up the direct interactions between solvent molecules and the mutating residue. In eq 8, we define a state function, which measures the changes in solvation due to the mutation. Indirect effects of the solvent on the structure and dynamics of the protein are automatically included.

We have performed the Met13 → Leu mutation in vacuo. $\Delta\Delta G$ in solution, from our calculations, is 1.6 kcal/mol. In vacuo, we find $\Delta\Delta G$ = 1.7 kcal/mol. The difference is related to the changes in the solvation associated with this mutation. From eq 8,

$$[G^{solv}(\text{ProPep}') - G^{solv}(\text{ProPep})] - [G^{solv}(\text{Pep}') - G^{solv}(\text{Pep})] = 0.1 \text{ kcal/mol}$$

Our calculation therefore predicts that the Met13 → Leu mutation affects the solvation of the peptide and the complex to the same extent. Therefore, the difference in affinity between the native and mutant peptides is not due to differences in solvation. It is due to interactions within the S-peptide and the S-protein.

DISCUSSION

The inaccuracies of our predicted structures limit the structural information we can draw from our simulations to a very qualitative level. On the other hand, the simulations probably give a reasonable idea of the importance of structural hysteresis and other errors in free energy calculations. We first discuss the various sources of error and then draw qualitative structural conclusions about S-protein–S-protein binding in RNase-S.

Table IV: Contribution of Structural Groups to the Free Energies[a]

	complex				peptide				ΔΔG			
contribution	total	cov	elec	vdW	total	cov	elec	vdW	total	cov	elec	vdW
					M → L Mutation[b]							
total	5.5	0.3	2.3	2.8	4.0	−1.4	2.0	3.3	1.6	1.7	0.3	−0.5
self[c]	1.2	0.3	−1.7	2.6	0.2	−1.4	−1.7	3.3	1.0	1.7	0.0	−0.7
interaction[d]	4.2		4.0	0.2	3.7		3.7	−0.6	0.5		0.3	0.8
solvent	2.1		2.0	0.1	2.7		2.4	0.3	−0.6		−0.4	0.2
protein	3.3	0.3	0.3	2.7	1.2	−1.4	−0.4	3.0	2.1	1.7	0.7	−0.3
Glu9	2.2		1.7	0.5	1.5		0.6	0.9	0.7		1.1	−0.4
Arg10	0.4		0.0	0.4	0.5		−0.1	0.6	−0.1		0.1	−0.2
Asp14	0.6		0.6	0.0	0.2		0.3	−0.1	0.4		0.3	0.1
Val47	0.7		0.4	0.3					0.7		0.4	0.3
Glu49	0.6		0.6	0.0					0.6		0.6	0.0
					I → L Mutation[e]							
total	−0.9	−1.1		0.2	−1.8	−0.9		−0.9	0.9	−0.2		1.1
self[c]	−0.9	−1.1		0.2	−1.0	−0.9		−0.1	0.1	−0.2		0.3
interaction[d]	0.0			0.0	−0.8			−0.8	0.8			0.8
solvent	0.6			0.6	−0.2			−0.2	0.8			0.8
protein	−1.5	−1.1		−0.4	−1.6	−0.9		−0.7	0.1	−0.2		0.3
Glu9	−0.5			−0.5	0.0			0.0	−0.5			−0.5
Arg10	0.0			0.0	0.3			0.3	−0.3			−0.3
His12	−0.4			−0.4	0.0			0.0	−0.4			−0.4
Asp14	0.4			0.4	0.0			0.0	0.4			0.4
Val47	0.5			0.5					0.5			0.5
					I → V Mutation[f]							
total	−2.0	0.1		−2.1	−2.0	−0.7		−1.3	0.1	0.8		−0.8
self[c]	−5.0	0.1		−5.1	−5.5	−0.7		−4.8	0.4	0.8		−0.3
interaction[d]	3.1			3.1	3.5			3.5	−0.4			−0.3
solvent	0.4			0.4	2.4			2.4	−2.0			−2.0
protein	−2.4	0.1		2.5	−4.4	−0.7		−3.7	2.0	0.8		1.2
Phe8	0.8			0.8	0.1			0.1	0.7			0.7
Glu9	0.7			0.7	0.3			0.3	0.4			0.4
Arg10	0.7			0.7	0.6			0.6	0.1			0.1
Val47	0.3			0.3					0.3			0.3
Ile51	0.3			0.3					0.3			0.3

[a] Results in are kcal/mol. [b] Averaged over one forward and one backward simulation. [c] Interactions within residue 13. [d] Interactions with the other residues and solvent. [e] Calculated from one I → L simulation. [f] Calculated from one I → V simulation.

The Multiple Minimum Problem and Sampling Errors. Free energy calculations are plagued by the multiple minimum problem. The deviations in the native RNase-S simulation from the X-ray structure, mainly at the end of the S-peptide, around His48, and at the protein surface, occur rapidly during the equilibration phase. They reflect inaccuracies in the empirical force field, limited sampling of conformation space, but also real differences between the crystal structure and the ensemble of solution structures. For example, the large variability of charged and polar surface side chains is to be expected, since these residues have significant experimental mobilities and positional error bars. Several of these side chains are modeled in more than one position in the X-ray structures (Kim et al., 1992; Varadarajan & Richards, 1992). The simulated mutants deviate still more from the X-ray structures. This is to be expected, since their prediction involves several specific problems. The simulations start from an initial X-ray structure and then must proceed over a free energy barrier of several kilocalories per mole, into a mutant structure with different side chain torsion angles and a different packing. The energy built up in the barrier passage enables the structure to rearrange, breaking helical hydrogen bonds and rotating side chains. This underlines a severe difficulty in structure prediction. When simply analyzing the energy of a known structure, we require that the force field give a reasonable energy estimate. To predict a new structure, however, the energy estimate must be accurate enough to distinguish the true energy well from numerous other energy wells/structures, which may differ by only a fraction of a kilocalorie per mole. The energy fluctuations in our tiny simulation system are themselves much greater than 1 kcal/mol. In addition, current force fields—including the one used here—usually neglect induced dipoles, nonpolar hydrogens, and various other effects. Furthermore, the energy barriers for rotation and buried dihedrals are large, limiting the structures that can be sampled and making it difficult for the system to relax into precisely the low-energy X-ray structure, even if the force field is able to distinguish it. Thus, our main hope for predicting reasonable mutant structures is in the small differences between the different mutant structures and the initial bias we achieve by starting from the appropriate X-ray structure. This was not sufficient in our case to converge to the mutant structures. Long simulations, which reduce random statistical errors, actually work against this initial bias.

An artificial cancellation of errors occurred here because we studied conformations of the RNase complex and of the S-peptide that were very similar. In the peptide simulations, the starting structure was taken directly from the RNase crystal structure. If a broader range of peptide conformations were used as starting points, the uncertainty associated with multiple conformations would increase.

In the future, we will be experimenting with umbrella potentials (Tobias & Brooks, 1987; Anderson & Hermans, 1988) that restrain the simulation toward the native and mutant crystal structures. In principle, some of the sampling problems could be alleviated with the use of massively parallel computers. In fact, the systematic exploration and simulation of a large number of local minima along the mutation pathway is well suited for parallel computation.

Other Errors in Free Energy Calculations. Force field limitations are another major source of error in these

Table V: Detailed Mutation Pathways

M → L
• 40-ps equilibration starting from the X-ray structure
• scale CG-SD-CE charges to zero (five linear steps; 10-ps equilibration + 30-ps data collection per step; coordinates saved every 0.04 ps)
• add CD2 atom to the structure with zero van der Waals weight and a CG–CD2 bond length of 0.3 Å
• change van der Waals parameters of CG and SD from Met values to Leu values (two-step thermodynamic perturbation calculation)
• turn off van der Waals interactions of CD2 and turn on those of CE; simultaneously modify geometrical parameters from Met values to Leu values (6 + 20-ps dynamics × 11 linear steps); CG–CD2 bond is increased to 1.526 Å; CE–SD bond is decreased to 0.3 Å
L → I
• 40-ps equilibration starting from the X-ray structure
• turn off CD2 1–4 van der Waals interactions (6 + 10-ps dynamics × 5 linear steps)
• shorten CG–CD2 bond to 0.3 Å (6 + 10-ps dynamics × 9 linear steps)
• turn off remaining CD2 van der Waals interactions (6 + 10-ps dynamics × 6 linear steps)
• delete CD2 and insert CG2 with CB–CG2 bond length of 0.3 Å
• change van der Waals parameters of CB, CG1, and CD1 from Leu values to Ile values in a two-step perturbation
• turn on CG2 van der Waals interactions other than 1–4 (6 + 10-ps dynamics × 6 linear steps)
• lengthen CB–CG2 bond to 1.526 Å (6 + 10-ps dynamics × 9 linear steps)
• turn on CG2 1–4 van der Waals interactions (6 + 10-ps dynamics × 5 linear steps)
I → V
• 40-ps equilibration starting from the X-ray structure
• turn off CD 1–4 van der Waals interactions (4 + 6-ps dynamics × 4 linear steps)
• shorten CG1–CD bond to 0.3 Å (4 + 6-ps dynamics × 7 linear steps)
• turn off remaining CD van der Waals interactions (4 + 6-ps dynamics × 6 linear steps)
• change van der Waals parameters of CG1 and CG2 from Ile values to Val values in a two-step perturbation

calculations. There is no easy way to alleviate these. One can only rely on the ongoing, gradual improvement of force fields, eventually including treatment of all hydrogens, induced dipoles, bulk solvent, and other refinements.

Random statistical errors are fairly easy to control, provided special care is taken for bond terms. All other terms have natural relaxation times that are long compared to the trajectory sampling. For the bond terms, it may be preferable to use fixed bond lengths, calculate the mean force along the bond, and derive the work for lengthening/shortening it (Tobias & Brooks, 1987; Anderson & Hermans, 1988).

Interpolation of the free energy derivative, estimated at a few discrete points, introduces a numerical error. Comparing the trapezoidal interpolation scheme to cubic spline interpolation, we find that the resulting differences in the free energy are small (~1–5%), except when the free energy derivative fluctuates strongly. This occurred only when the bond term fluctuated spuriously, due to a bad choice of sampling frequency.

Error asociated with the origin singularity for atom insertion was not a factor in this work, since we introduced/deleted atoms with artificially shortened bonds. These errors can potentially be rather large (Simonson, unpublished results). It is probably best either to proceed as here, with bond shortening, or to use at least a λ^{12} scaling of the appropriate van der Waals term. This can be done by scaling the van der Waals radius of the appearing particle in a linear way.

Structural Contributions to the Binding Thermodynamics in RNase-S. Despite the inaccuracy of the predicted mutant structures, our simulations suggest several structural conclusions. First, it appears that several packing arrangements are available to the protein, with only modest energy differences between them. For example, in the Met13 → Leu simulation, the χ_1 and χ_2 angles of residue 13 remain in their native values throughout the simulation, while the M13L X-ray structure has different values. These different packing arrangements lead to somewhat different values of ΔG(Com) and ΔG(Pep). Using the same starting conformations for the S-peptide in the complex and in solution, these differences tended to cancel out, giving similar $\Delta\Delta G$'s for the different calculations. The possibility of several packing arrangements is consistent the crystal structures (Kim et al., 1992; Varadarajan & Richards, 1992), where several internal side chains are modeled in two positions.

The partial changes on the Met13 side chain are modest: C_γ, 0.235 au; S_δ, –0.47 au; C_ϵ, 0.235 au. In a survey of 307 methionines in high-resolution protein structures, only 25% of the sulfurs were within 4 Å of potential hydrogen-bonding partners (Gregoret et al., 1991). In RNase-S, the Met13 side chain is buried and not involved in a hydrogen bond. Nevertheless, its electrostatic interaction energy with its surroundings is significant: –5 kcal/mol on average in our native simulation. The free energy of removing the partial charges is estimated to be about 2 kcal/mol, both in the RNase-S complex and in the S-peptide in solution. The groups that contribute to this free energy are mainly Glu9, His12, Asp14, the Met13 backbone, and the solvent. The solvation of these charges by the protein and the peptide are equivalent; therefore, the charges do not stabilize or destabilize the complex with respect to the unbound state.

The free energies obtained here result from small, direct interactions with a few residues in the immediate vicinity of residue 13. One or two solvent molecules are also close enough to contribute van der Waals interactions at any given moment in the simulations. These small van der Waals interactions tend to cancel when the complex and peptide are compared, giving still smaller $\Delta\Delta G$ contributions. The most important single group contributing to the free energy is residue 13 itself, through 1–4 and 1–5 repulsive van der Waals interactions between the side chain and the backbone. This is partly because of the bond shortening/lengthening protocol used here. There is no simple interpretation of the free energies in terms, say, of protein–peptide repulsive contacts in the complex being replaced by soft solvent contacts in solution.

CONCLUSIONS

In liquid simulations, a good criterion for precision of a calculation is the reduction of the random statistical error (Schiferl & Wallace, 1985; Straatsma et al., 1986). However, in free energy calculations on proteins (Tropscha & Hermans, 1992), this criterion is not sufficient. Indeed, there exist a finite number of important local energy minima. In

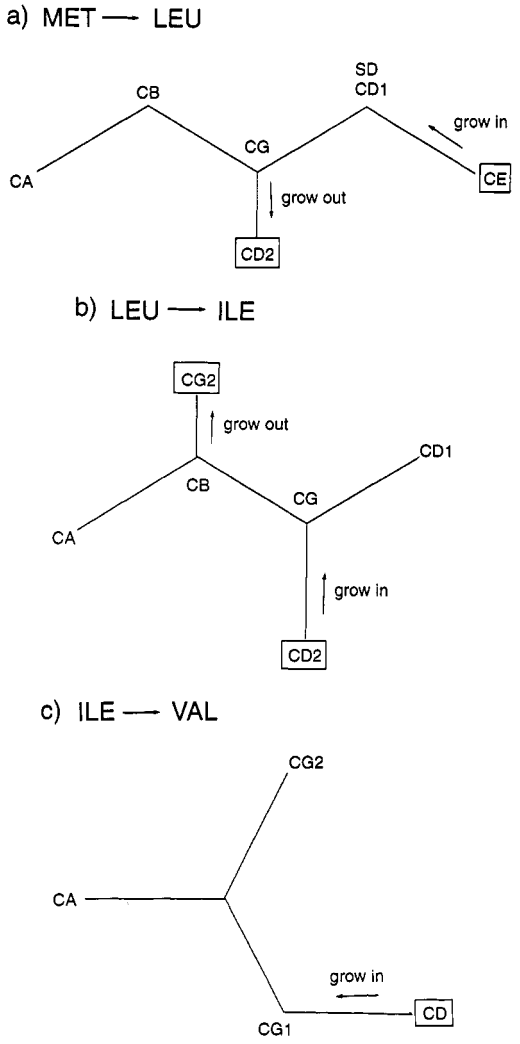

FIGURE 8: Schematic view of the hybrid amino acids involved in the various mutations. (a) Met → Leu; (b) Leu → Ile; (c) Ile → Val. Changes due to differences in geometry and partial charges are not indicated. Details are given in Table V.

our case, these corresponded to different sets of dihedral rotamers for the residues around position 13. If just one or two minima are visited in a simulation, the statistical error will become small after just a few picoseconds of sampling. But if a second simulation happens to sample different local minima, the calculated free energy will be different, albeit with an equally small apparent statistical error. We obtained random statistical errors of 0.5–2 kcal/mol, but with more careful bond sampling these could be reduced considerably. The presence of multiple structures introduced an additional uncertainty of 1–2 kcal/mol. If more simulations of each mutation were run, the uncertainty associated with multiple structures would increase. In general, the apparent statistical error can be made quite small in any given simulation. However, the uncertainty due to multiple structures is much greater and more difficult to control (Hodel et al., 1992). Thus, a better criterion for precision of a protein free energy simulation is that one must systematically identify and visit the entire set, hopefully small, of important local minima and reduce the statistical error associated with each visit.

In future work, it will be necessary to bias the simulations more toward the known experimental structures, by using some form of restraining umbrella potential, for example (Beveridge & di Capua, 1989). The mutation pathway must be designed with great care to reduce the energy built up in the barrier passage. Some energy terms can be reduced in certain parts of the simulation. Reduction of energy barriers will enhance the sampling of dihedral torsion angles. With a very specific selection of energy terms, it should be possible to converge to the final experimental structure. In general, it appears more important to do a multitude of short simulations with a smart, user-directed sampling, rather than very long simulations that rely purely on random sampling.

ACKNOWLEDGMENT

We acknowledge helpful discussions with A. Hodel, E. Kim, M. Nilges, F. Richards, H. Treutlein, R. Varadarajan, and H. Wyckoff.

APPENDIX: COMPLETE LIST OF MUTATION PATHWAYS

The exact pathways followed in each simulation are described in Table V. For each entry a backward simulation was also done, starting from the appropriate X-ray structure and following the reverse pathway. All steps used thermodynamic integration except where specified (the van der Waals parameter changes). In each mutation the minimum number of atoms is used; for example, in the Met → Leu case, Met-SD is mapped into Leu-CD1, while Met-CE is removed and Leu-CD2 is grown in (see Figure 8). The number of steps and the amount of dynamics per step vary somewhat from one mutation to the next. Overall, we were very generous for the M → L simulation and less so for L → I and I → V.

REFERENCES

Alber, T. (1989) Mutational effects on protein stability, *Annu. Rev. Biochem. 58*, 765–798.

Anderson, A., & Hermans, J. (1988) Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamics simulations, *Proteins 3*, 262–265.

Bash, P., Singh, U. C., Brown, F., Langridge, R., & Kollman, P. (1987) Calculations of the relative change in binding free energy of a protein–inhibitor complex, *Science 235*, 574–576.

Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., & Haak, J. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 811*, 3684–3690.

Beveridge, D., & DiCapua, F. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems, *Annu. Rev. Biophys. Biophys. Chem. 18*, 431–492.

Blackburn, P., & Moore, S. (1982) Pancreatic ribonuclease, in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. XV, pp 317–433, Academic Press, New York.

Borah, B., Chen, C. W., Egan, W., Miller, M., Wlodawer, A., & Cohen, C. W. (1985) Nuclear magnetic resonance and neutron diffraction studies of the complex of ribonuclease A with uridine vanadate, a transition state analogue, *Biochemistry 24*, 2058–2067.

Brooks, C. (1988) Thermodynamic calculations on biological molecules, *Int. J. Quant. Chem. Quant. Biol. Symp. 15*, 221–234.

Brooks, C., & Karplus, M. (1984) Deformable stochastic boundaries in molecular dynamics, *J. Chem. Phys. 79*, 6312–6325.

Brooks, C., Brünger, A., & Karplus, M. (1985) Active site dynamics in proteins: a stochastic boundary molecular dynamics approach; *Biopolymers 24*, 843–865.

Brünger, A. T. (1987) X-PLOR Version 2.1, Yale University, New Haven, CT.

Brünger, A., Brooks, C., & Karplus, M. (1985) Active site dynamics of ribonuclease, *Proc. Natl. Acad. Sci. U.S.A. 82*, 8458–8462.

Campbell, R., & Petsko, G. (1987) Ribonuclease structure and catalysis: crystal structure of sulfate-free native ribonuclease A at 1.5 Å resolution, *Biochemistry 26*, 8579–8584.

Connelly, P., Varadarajan, R., Sturtevant, J., & Richards, F. (1990) Thermodynamics of protein–peptide interactions in the ribonuclease-S system studied by titration calorimetry, *Biochemistry 29*, 6108–6114.

Creighton, T. (1991) Stability of folded conformations, *Curr. Opin. Struct. Biol. 1*, 5–16.

Dang, L., Merz, K., Jr., & Kollman, P. (1989) Free energy calculations on protein stability: Thr157 to Val157 mutation of T4 lysozyme, *J. Am. Chem. Soc. 111*, 8505–8508.

Davenport, W., Jr. (1970) *Probability and Random Processes*, McGraw-Hill, New York.

Gao, J., Kuczera, K., Tidor, B., & Karplus, M. (1989) Hidden thermodynamics of mutant proteins: a molecular dynamics analysis, *Science 244*, 1069–1072.

Gregoret, L., Rader, S., Fletterick, R., & Cohen, F. (1991) Hydrogen bonds involving sulfur atoms in proteins, *Proteins 9*, 99–107.

Griffin, J., Schechter, A., & Cohen, J. (1973) Nuclear magnetic resonance studies of a ribonuclease–dinucleoside phosphonate complex and their implications for the mechanism of the enzyme, *Ann. N.Y. Acad. Sci. 222*, 693–707.

Haar, W., Maurer, W., & Rüterjans, H. (1974) Proton magnetic resonance studies of complexes of pancreatic ribonuclease A with pyrimidine and purine nucleotides, *Eur. J. Biochem. 44*, 201–251.

Hodel, A., Simonson, T., & Brünger, A. T. (1992) Conformational substates and uncertainty in free energy calculations (submitted).

Howlin, B., Moss, D., & Harris, G. (1990) Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model, *Acta Crystallogr. A45*, 851–861.

Janin, J., Wodak, S., Levitt, M., & Maigret, B. (1978) Conformations of amino acid side chains in proteins, *J. Mol. Biol. 125*, 357–386.

Jorgensen, W. (1989) Free energy calculations: a breakthrough for modelling organic chemistry in solution, *Acc. Chem. Res. 22*, 184–189.

Jorgensen, W., & Tirado-Rives, J. (1988) The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc. 110*, 1657–1666.

Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R., & Klein, M. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926–935.

Kellis, J., Nyberg, K., & Fersht, A. (1989) Energetics of complementary side chain packing in a protein hydrophobic core, *Biochemistry 28*, 4914–4922.

Kim, E., Varadarajan, R., Wyckoff, H., & Richards, F. (1992) Refinement of the crystal structure of ribonuclease S. Comparison with and between the various ribonuclease A structures, *Biochemistry* (in press).

Kirkwood, J. (1935) Statistical mechanics of fluid mixtures, *J. Chem. Phys. 3*, 300–313.

Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946–950.

Lenstra, J., Bolscher, B., Stob, S., Beintema, J., & Kaptein, R. (1979) The aromatic residues of bovine pancreatic ribonuclease studied by ^{1}H nuclear magnetic resonance, *Eur. J. Biochem. 98*, 385–397.

Lim, W., & Sauer, R. (1989) The role of internal packing interactions in determining the structure and stability of a protein, *J. Mol. Biol. 219*, 359–376.

Markley (1976) Correlation proton magnetic resonance studies at 250 MHz of bovine pancreatic ribonuclease, *Biochemistry 14*, 3546–3566.

Matthews, B. (1991) Mutational analysis of protein stability, *Curr. Opin. Struct. Biol. 1*, 17–19.

Matthews, B., et al. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect, *Science 225*, 178–183.

McQuarrie, D. (1976) *Statistical Mechanics*, Harper and Row, New York.

Meadows, D., Jardetzky, O., Epand, R., Rüterjans, H., & Scheraga, H. (1969) Assignment of the histidine peaks in the nuclear magnetic resonance spectrum of ribonuclease, *Proc. Natl. Acad. Sci. U.S.A. 60*, 766–772.

Niu, C. H., Matsuura, S., Shindo, H., & Cohen, J. S. (1979) Specific peptide–protein interactions in the ribonuclease S′ system studied by ^{13}C nuclear magnetic spectroscopy with selectivity ^{13}C-enriched peptides, *J. Biol. Chem. 254*, 3788–3796.

Pearlman, D., & Kollman, P. (1989) A new method for carrying out free energy calculations: dynamically modified windows, *J. Chem. Phys. 90*, 2460–2470.

Pomes, R., & McCammon, J. A. (1990) Mass and step length optimization for the calculation of equilibrium properties by molecular dynamics simulation, *Chem. Phys. Lett. 166*, 425–428.

Ponder, J., & Richards, F. (1987) Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol. 193*, 775–791.

Postma, J., Berendsen, H., & Haak, J. (1982) Thermodynamics of cavity formation in water. A molecular dynamics study, *Faraday Symp. Chem. Soc. 17*, 55–67.

Prévost, M., Tidor, B., Karplus, M., & Wodak, S. (1991) Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile96 → Ala mutation in barnase, *Proc. Natl. Acad. Sci. U.S.A. 88*, 10880–10884.

Richards, F. M. (1955) On an active intermediate produced during the digestion of ribonuclease by subtilisin, *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 329–346.

Richards, F. M. (1977) Areas, volumes, packing, and protein structure, *Annu. Rev. Biophys. Bioeng. 6*, 151–175.

Richards, F. M., & Wyckoff, H. W. (1971) Bovine pancreatic ribonuclease, in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. IV, pp 647–907, Academic Press, New York.

Richards, F. M., & Wyckoff, H. W. (1973) *Ribonuclease S, Atlas of Molecular Structures in Biology* (Phillips, D. C., & Richards, F. M., Eds.) Vol. I, Clarendon Press, Oxford.

Rico, M., Bruix, M., Santoro, J., Gonzalez, C., Neira, J. L., Nieto, J. L., & Herrantz, J. (1989) Sequential ^{1}H-NMR assignment and solution structure of bovine pancreatic ribonuclease A, *Eur. J. Biochem. 183*, 623–638.

Rico, M., Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., Nieto, J. L., & Herrantz, J. (1991) 3D structure of bovine pancreatic ribonuclease A in aqueous solution: an approach to tertiary structure determination from a small basis of ^{1}H-NMR NOE correlations, *J. Biomol. NMR 1*, 283–298.

Robertson, A., Purisma, E., Eastman, M., & Scheraga, H. (1989) Proton NMR assignments and regular backbone structure of bovine pancreatic ribonuclease S in aqueous solution, *Biochemistry 28*, 5930–5938.

Rüterjans, H., & Witzel, H. (1969) NMR studies on the structure of the active site of pancreatic ribonuclease A, *Eur. J. Biochem. 9*, 118–127.

Ryckaert, J., Ciccotti, G., & Berendsen, H. (1977) Numerical integration of the Cartesian equations of motion for a system with constraints: molecular dynamics of *n*-alkanes, *J. Comput. Phys. 23*, 327–341.

Sandberg, W., & Terwilliger, T. (1989) Influence of interior packing and hydrophobicity on the stability of a protein, *Science 245*, 54–57.

Schiferl, S., & Wallace, D. (1985) Statistical errors in molecular dynamics averages, *J. Chem. Phys. 83*, 5203–5209.

Shortle, D., Stites, W., & Meeker, A. (1990) Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease, *Biochemistry 29*, 8033–8041.

Straatsma, T., Berendsen, H., & Stam, A. (1986) Estimation of the statistical error in molecular simulation calculations, *Mol. Phys. 57*, 89–95.

Swaminathan, S., Ravishanker, G., Beveridge, D., Lavery, R., Etchebest, C., & Sklenar, H. (1990) Conformational and helicoidal analysis of the molecular dynamics of proteins: "Curves", dials and windows for a 50 ps dynamic trajectory of BPTI, *Proteins 8*, 179–193.

Taylor, H., Richardson, D., Richardson, J., Wlodawer, A., Komoriya, A., & Chiaken, I. (1981) "Active" conformation of an inactive semi-synthetic ribonuclease S, *J. Mol. Biol. 149*, 313–317.

Tembe, B., & McCammon, J. A. (1984) Ligand–receptor interactions, *Comput. Chem. 8*, 281–283.

Tidor, B., & Karplus, M. (1991) Simulation analysis of the stability mutant R96H of T4 lysozyme, *Biochemistry 30*, 3217–3228.

Tobias, D., & Brooks, C. L., III (1987) Calculation of free energy surfaces using the methods of thermodynamic perturbation theory, *Chem. Phys. Lett. 142*, 472–476.

Tropscha, A., & Hermans, J. (1992) Application of free energy simulations to the binding of a transition-state analogue inhibitor to HIV protease, *Protein Eng. 5*, 29–33.

van Gunsteren, W., & Weiner, P. (1989) *Computation of Free Energy for Biomolecular Systems*, Escom Science Publishers, Leiden.

Varadarajan, R., & Richards, F. (1992) Crystallographic structures of ribonuclease S variants with non-polar substitutions at position 13: packing and cavities, *Biochemistry* (in press).

Varadarajan, R., Connelly, P., Sturtevant, J., & Richards, F. (1992) Heat capacity changes for protein-peptide interactions in the ribonuclease-S system, *Biochemistry 31*, 1421–1426.

Verlet, L. (1967) Computer "experiments" on classical fluids. I. Thermodynamic properties of Lennard-Jones molecules, *Phys. Rev. 159*, 98.

Walters, D., & Allerhand, A. (1980) Tautomeric states of the histidine residues of bovine pancreatic ribonuclease A. Application of carbon 13 nuclear magnetic spectroscopy, *J. Biol. Chem. 255*, 6200–6204.

Warshel, A., & Aqvist, J. (1989) Microscopic simulations of chemical processes in proteins and the role of electrostatic free energy, in *Theoretical Biochemistry and Molecular Biophysics* (Beveridge, D. L., & Lavery, R., Eds.) pp 257–273, Adenine Press, New York.

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

Witzel, H. (1963) The function of the pyrimidine base in the ribonuclease reaction, *Progr. Nucleic Acid Res. 2*, 221–258.

Wlodawer, A., Bott, R., & Sjolin, L. (1982) The refined structure of ribonuclease A at 2 Å resolution, *J. Biol. Chem. 257*, 1325–1332.

Wong, C., & McCammon, J. A. (1986) Dynamics and design of enzymes and inhibitors, *J. Am. Chem. Soc. 108*, 3830–3832.

Wyckoff, H., Hardman, K., Allewell, N., Inagami, T., Johnson, L., & Richards, F. (1967) The structure of ribonuclease S at 3.5 Å resolution, *J. Biol. Chem. 242*, 3984–3988.

Wyckoff, H., Tsernoglou, D., Hanson, A., Knox, J., Lee, B., & Richards, F. (1970) The 3-dimensional structure of ribonuclease S. *J. Biol. Chem. 245*, 305–328.

Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) Dependence of conformation stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase α-subunit, *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441–4444.

Zwanzig, F. (1954) High-temperature equation of state by a perturbation method. I. Non-polar gases, *J. Chem. Phys. 22*, 1420–1426.