A Procedure for Detection and Quantitation of Cavity Volumes in Proteins

APPLICATION TO MEASURE THE STRENGTH OF THE HYDROPHOBIC DRIVING FORCE IN PROTEIN FOLDING*

Received for publication, February 11, 2002, and in revised form, June 13, 2002
Published, JBC Papers in Press, June 17, 2002, DOI 10.1074/jbc.M201373200

Suvobrata Chakravarty‡, Akshay Bhinge‡, and Raghavan Varadarajan¶¶

From the ‡Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012 and the ¶¶Chemical Biology Unit, Jawaharlal Center for Advanced Scientific Research, Jakkur P. O., Bangalore 560 004, India

Accurate identification of cavities is important in the study of protein structure, stability, design, and ligand binding. Identification and quantitation of cavities is a nontrivial problem because most cavities are connected to the protein exterior. We describe a computational procedure for quantitating cavity volumes and apply this to derive an estimate of the hydrophobic driving force in protein folding. A grid-based Monte Carlo procedure is used to position water molecules on the surface of a protein. A Voronoi procedure is used to identify and quantify empty space within the solvated protein. Additional cavities not detected by other existing procedures can be identified. Most of these are close to surface concavities. Residue volumes for both the interior and the surface residues as well as cavity volumes are in good agreement with volumes calculated from fully hydrated protein structures obtained from molecular dynamic simulations. We show that the loss of stability because of cavity-creating mutations correlates better with cavity volumes determined by this procedure than with cavity volumes determined by other methods. Available structural and thermodynamic data for a number of cavity-containing mutants were analyzed to obtain estimates of 26.1 cal·mol⁻¹·Å⁻³ and 18.5 cal·mol⁻¹·Å⁻² for the relative contributions of cavity formation and the hydrophobic effect to the observed stability changes. The present estimate for the hydrophobic driving force is at the lower end of estimates derived from model compound studies and considerably lower than previous estimates of approximately 50 cal·mol⁻¹·Å⁻² derived from protein mutational data.

In the absence of structural rearrangement, on average, deletion of a single methylene group is expected to result in losses in stability of 0.41 and 0.70 kcal·mol⁻¹ resulting from decrease in hydrophobicity and packing, respectively.

An understanding of the folding and stability of proteins depends on an accurate evaluation of the interaction of buried groups with each other and the interaction of exposed groups with solvents. The basic concepts of packing density, packing interaction, atomic volume, and solvent-accessible surface area have been extensively applied to such studies for many years (1–4). These studies have shown that close packing is an important feature of protein structure and stability. While studying the packing density of lysozyme and ribonuclease, it was determined that the average value of packing density in the protein interior is close to 0.75 (1). Although on average the protein interior is highly packed, there are variations in the packing efficiency. In bovine pancreatic ribonuclease (RNase A), for example, this number ranges from 0.66 to 0.84 in different regions of the protein (1). Cavities in proteins are a major contributor to low packing densities and decreased stability (5). Cavities and surface grooves are also potential sites for binding of drugs, ligands, and other proteins. Hence, identification and detection of cavities is important in the study of protein structure, stability, and design.

There are several existing procedures/programs such as VOIODOO (6), MS package (7, 8), VOLBL (9, 10), and CAST (11) for cavity detection in proteins. However, some of them are often unable to detect cavities close to the protein surface or to identify surface grooves. We present here a simple and robust procedure to detect cavities in proteins. The method uses a Monte Carlo (MC)1 procedure described previously (12) to solvate the protein and to identify internal cavities. We compare the performances of the existing programs with that of the MC method here.

For an infinite set of arbitrary points/atoms in space, a geometrical procedure called the Voronoi procedure (13) divides up space with a unique volume assigned to each point or atom. Voronoi volume calculations are one of the best methods available to study packing in proteins (1). To determine the volume of space associated with an atom, the procedure requires the location of all neighbors. The procedure works well for atoms in the protein core. However, for protein atoms on the surface, many of the neighbors are often disordered water molecules that are not detected in protein crystal structures or nuclear magnetic resonance (NMR) structures (14, 15). To address this problem, using bovine pancreatic trypsin inhibitor as an example, molecular dynamic (MD) simulations have been used (16) to realistically position water molecules around the protein. Such simulations are in principle the most appropriate method to overcome the difficulty in calculation of surface residue volume using the Voronoi procedure but are computa-

* This work was supported by grants from Department of Science & Technology (India) and Department of Biotechnology (India) (to R. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Recipient of a Swarnajayanthi fellowship from the Government of India and senior research fellow of the Wellcome Trust. To whom correspondence should be addressed. Fax: 91-80-3600535 or 91-80-3600683; E-mail: varadar@nbus.ernet.in.

†† The abbreviations used are: MC, Monte Carlo; RCSB, Research Collaboratory for Structural Bioinformatics; SNase, staphylococcal nuclease; MD, molecular dynamic; ASA, accessible surface area; TIP3P, transferable intermolecular potential.
tionally very expensive. We make use of a simple procedure described in earlier work (12) to overcome the problem of calculating volumes of surface protein atoms. The results of our procedure are in good agreement with volumes calculated from MD simulations.

It is generally thought that the hydrophobic effect is one of the major factors in stabilizing folded proteins (17). However, there has been considerable debate regarding the quantitative contribution of the hydrophobic effect to protein stability. Solubility and vapor pressure measurements on nonpolar solutes indicate a value of 25–30 cal·mol⁻¹·Å⁻² (18–20). However, surface tension measurements at a hydrocarbon water interface suggest a value as high as 72 cal·mol⁻¹·Å⁻². It was suggested that effects of curvature, surface roughness, and disorder account for this apparent difference and that a revised value for the strength of the hydrophobic effect is 47 cal·mol⁻¹·Å⁻² (4, 21, 22). This value was also in agreement with results of some protein engineering studies (23). In a typical protein engineering experiment, a hydrophobic residue within the core of a protein is substituted by a smaller hydrophobic residue. The resulting change in the stability is taken as measure of the difference between the hydrophobic stabilization provided by the two amino acids. However, such experiments carried out with different proteins or at different sites in the same protein have given varied results (23–27). This is possibly because of the differences in local packing interactions and cavity formation, which were not taken into account. A recent analysis of protein mutational data and small molecule transfer studies highlighted some significant differences between the two sets of data (28). It was suggested that the strength of the hydrophobic driving force was likely to be somewhat lower than the estimates of 25–30 cal·mol⁻¹·Å⁻² derived from the transfer studies. Other work has concluded that the unfavorable transfer free energies of nonpolar molecules from nonpolar solvents to water are primarily because of more favorable dispersion interactions in the nonpolar solvent, rather than an unfavorable free energy of interaction between nonpolar solute and water in the aqueous phase (29). In the present work we have tried to quantify the relative strengths of packing and hydrophobic interactions in a number of cavity-containing mutants for which both structural and thermodynamic data were available. The data were analyzed to obtain relative estimates of the contributions of cavity formation and the hydrophobic effect to the observed stability changes in proteins.

MATERIALS AND METHODS

Proteins Used for Volume Calculations—All the protein structure coordinates were taken from the Research Collaboratory for Structural Bioinformatics (RCSB) (48). The following RCSB identification codes were used in the study of model mutants. The residues mutated to Ala are indicated in parentheses. 193L (Ile-55, Leu-56, Met-12, Trp-28), 1A19 (Phe-74, Leu-51, Trp-53), 1BVI (Phe-80, Trp-59), 1DFR (Phe-125, Met-42, Met-92), 1EQL (Phe-57, Met-101, Val-24, Val-63), 1HFX (Ile-30, Ile-55, Ile-56, Leu-2, Met-29, Trp-28), 1MBD (Met-131), 1RUV (Phe-46, Phe-8, Ile-106, Met-42, Val-9), 1STN (Phe-34, Ile-92, Leu-103, Val-104, Val-23), 2A2P (Ile-88, Leu-89), 2CT2 (Ile-92), 5ICB (Leu-52, Val-61). The following RCSB IDs were used in the study of energetics of cavity formation. RNase S (1RNV, 1BRG, 1RBD, 1RBF, 1RBI, 1RBC, 1RHC), T4 lysozyme (1L63, 1L67, 1L69, 1L59, 2001, 235L, 236L, 238L, 240L, 242L, human lysozyme (1LZ1, 1IYA, 1YAO, 1YAP, 1IYQ, 1UB, 1UC, 1UD, 1UE, 1UF, 1UG, 1UH, 1ULJ, 1ULJ, 1BR1, 1BR2, 1BR3, 1BR4, 1BR5, 1BR6, 1BR7), and Gene V protein (1GV, 1VQ1, 1VQA).

Voronoi Program—The Voronoi program was obtained from bioinfo.mbb.yale.edu/geometry/beta, ftp://hyper.stanford.edu/pub/mbg (16, 30).

MC Cavitation Calculations—Cavities and the Hydrophobic Effect

To obtain volumes of cavities close to the site of a large to small substitution, the wild type protein was superposed on the mutant protein (31) using only the Ca atoms. The number of internal water molecules, \(N_m\) and \(N'_m\) (for the mutant and the wild type, respectively), in a sphere of radius 5.5 Å from the side-chain centroid of the wild type residue were counted. The difference between the sum of the Voronoi volumes of the \(N_m\) and \(N'_m\) internal waters corresponded to the cavity volume \(\triangle \text{v} \text{cavity volume} \). This distance of 5.5 Å was chosen because, in the examples examined in this work, additional cavities (present only in the mutant protein but not in the wild type) of sufficient size to accommodate a water molecule were found only in the vicinity of the mutation. All appreciable structural changes were also local. It is desirable to have a cut-off to avoid considering artificial small cavities and surface grooves that occur all over the protein. Small differences in the coordinates of wild type and mutants that are within the coordinate error sometimes result in differences in the locations and values of such cavity volumes. It is very difficult to evaluate the cumulative energetic effects of small, distributed changes in coordinates, but the excellent correlation observed (see below) between changes in stability and observed cavity volumes suggests that, in the systems examined, these effects are not appreciable.

Cavity Detection by VOIDOO, MS, and CAST—Cavity volumes were computed using VOIDOO (6), MS package (8), and CAST (11) and compared with those of the MC method. The VOIDOO program was obtained from x-ray.bmc.uu.se/as/VOIDOO.html. For all our comparisons, a probe radius of 1.2 Å was used for VOIDOO and the MS package. This radius was shown previously to be optimal for cavity detection by the MS program (32). Each of these cavity detection programs measures different kinds of cavity volumes so the numbers are not directly comparable. CAST uses an alpha shape-based procedure to measure the cavity enclosed by the molecular surface of atoms that surround the cavity. VOIDOO is a grid-based procedure that measures the cavity volume defined by the van der Waals surface of atoms lining the cavity. Both CAST and VOIDOO have options to measure cavities enclosed by the solvent-accessible surface, but we did not use this. The MS program measures cavity volumes accessible to a probe sphere enclosed by the molecular surface of cavity lining atoms. The MC procedure developed in the present work measures the Voronoi volume of a cluster of overlapping spheres that map the cavity. The MS program reports the center of the cavity, whereas CAST reports only the atoms in the boundary of the cavity. The center of a cavity detected by CAST was generated from the center of mass of the boundary atoms. The centers of the cavities were represented by water molecule for visual examination using computer graphics. The results of CAST were received from cast.scriibo.edu/cgi-bin/cast1/index.pl (11). For all the CAST calculations, the server default radius of 1.4 Å was used. In all discussions above, the volume of the cavity created because of large to small residue substitutions was obtained from the difference between the cavity volume in the mutant and the wild type protein (Δcavity volume). Only cavity volumes close to the site of mutation were selected, and the total cavity volume of these selected cavities was calculated. To compare cavity volumes calculated by different methods, model mutants were constructed and the cavity volumes calculated by each method were normalized as described under "Model Mutants."

Δ\(\Delta G\) and ASA Calculations—The loss of stability in cavity-containing mutants, \(\Delta G_{\text{folded}}^{\text{mutant}} - \Delta G_{\text{folded}}^{\text{wild type}}\) is obtained as follows.

\[
\Delta G_{\text{folded}}^{\text{mutant}} = G_{\text{folded state}}^{\text{mutant}} - G_{\text{folded state}}^{\text{wild type}}
\]

\[
\Delta G_{\text{folded}}^{\text{mutant}} = G_{\text{folded state}}^{\text{mutant}} - G_{\text{folded state}}^{\text{wild type}}
\]

Solvent-accessible surface area (ASA) was calculated using the implementation of the Lee and Richards (33) algorithm with a probe radius of 1.4 Å and a z-section of 0.05Å. Solvent accessibility of a residue X was calculated as the ratio of the observed ASA of X in the folded state to its ASA in a Gly-X-Gly peptide of extended conformation (18). The Δ\(\Delta\)ASA resulting from mutation of a particular residue was determined as follows.

\[
\Delta \text{ASA} = \text{ASA}(\text{folded state}) - \text{ASA}(\text{unfolded state/Gly-X-Gly})
\]

\[
\Delta \text{ASA} = \Delta \text{ASA}(\text{mutant}) - \Delta \text{ASA}(\text{wild type})
\]
**RESULTS AND DISCUSSION**

**Procedure for Volume Calculations by MC Method**—The protein molecule of interest is placed at the center of a pre-equilibrated box of transferable intermolecular potential (TIP3P) water obtained from a Monte Carlo simulation. Different configurations of solvent ensembles around the protein are generated in an iterative fashion as described previously (12). In every iteration, water molecules that sterically overlap with protein are removed, leaving internal water molecules (those present in the cavities and surface grooves) and bulk water molecules (those surrounding the protein) (12). Shown in Fig. 1 is a schematic of a protein structure depicting positions of water molecules in the internal cavity, C, and surface groove, G, of a protein along with other hypothetical water molecules detected in a single iteration. At the end of several iterations (usually 25), internal water molecules detected in grooves and cavities are clustered into distinct groups. An empty space at which a cluster of internal water molecules is detected is considered to be a cavity or a surface pocket. The Voronoi volume of these clusters is calculated to obtain an estimate of the cavity volume. The Voronoi volume calculations are also performed on a solvated protein without and with the cluster of internal waters. Residue volumes calculated in the presence and absence of the cluster of internal waters give us an estimate of the excess space associated with residues that surround cavities.

Because of the empirical nature of the above procedure, there are some inherent limitations. The minimum size of detected cavities is on the order of ~9 Å³, close to the van der Waals volume of water molecule ~10 Å³. The maximum size of a spherical cavity detected in this procedure is ~310 Å³, the size of a sphere of radius 4.2 Å. However, most large cavities in proteins are nonspherical in nature, so this is not a serious limitation.

**Calculation of Residue Volumes**—After removal of internal water molecules in each iteration, volumes of protein atoms in the solvated protein structures were calculated using the Voronoi procedure (1, 13, 16, 30). The Voronoi polyhedra were constructed using the radial plane method (35). Residue volumes were determined from the sum of the constituent atom volumes. In all our calculations, 25 iterations were used to generate different ensembles of surface water molecules. The volume of a residue is taken as the average of its volumes calculated in 25 iterations. To check the accuracy of the method, a comparison was made with similar calculations performed with solvated protein structures obtained during the course of MD simulations. Structures were extracted at different time points in the course of simulations. The simulations were carried out on RNase S (124 residues, 1RNV), barnase (110 residues, 1BNI), and T4 lysozyme (162 residues, 1L63) solvated by 2443, 2248, and 3096 water molecules, respectively. A comparison of the residue volumes calculated using structures from MD simulations with those calculated using the MC procedure is shown in Fig. 2. The results show the percentage of change in residue volume calculated using our procedure with respect to that obtained using structures from MD simulations as a function of residue depth (12). Residue volumes calculated with our procedure are in good agreement with those from MD simulations, although surface residue volumes are slightly higher using our MC method (Fig. 2). Approximately 25% of residues show 10% or higher volume in the MC method. The average depth of residues showing 10% or higher volume is 4.2 Å, indicating that only surface residues show this increase in the MC method. During the course of MD simulation, some solvent water molecules are able to come close to the protein surface atoms because of highly directional, specific charge and dipolar interaction. This results in slightly smaller Voronoi polyhedra than those constructed in our approach, which uses only a simple distance-based criterion for solvent positioning.

**Model Mutants**—Because each of the cavity volume procedures (CAST, VOIDOO, MS, and MC) computes different measures of the cavity volume, it is necessary to normalize the calculated values to allow comparisons. To carry out such a normalization as well as to compare the accuracy of volumes determined by the various procedures, we carried out calculations on artificial/model mutants of 14 proteins where large residues were replaced with smaller residues. Model mutants were created by deleting the side-chain atoms of the selected protein molecule.
residue from the Protein Data Bank structure file. For example in a Phe → Ala model mutant, all Phe side-chain atoms after the CB atom would be deleted. We constructed a total of 43 X → Ala model mutants where X stands for Trp, Phe, Met, Leu, Ile, and Val. Only bulky residues with side-chain depth 8.0 Å were chosen. This was done to ensure proper comparison with the expected volumes as Voronoi volumes of only buried (18, 19, 30) residues are accurately known. Fig. 3 shows comparisons of cavity volumes determined by the various methods. The cavity volumes determined by all four methods were compared with the side-chain Voronoi volume of Ala from that of residue X (19). The average expected cavity volumes of X → Ala model mutants were found to be related to the calculated volumes by linear relationships in all four cases (Fig. 3, Table I). This indicates that all the methods overestimate cavity volume at buried sites in a consistent and correctable fashion. In the MC method, this is probably the result of the use of a uniform cut-off distance of 2.6 Å to select the internal waters instead of using atom-specific van der Waals radius for atoms. The actual Voronoi volume of a particular cavity can be obtained using the following corrections, where 

\[ \text{Expected Voronoi volume} = m_1 |\Delta \text{cavity volume}| + m_2 \]  

(5)

Unless otherwise stated, all cavity volumes reported hereafter are corrected using the above equations with values of \( m_1 \) and \( m_2 \) listed in Table I. Normalizing calculated volumes to Voronoi volumes is useful for many applications, as the Voronoi volume is the closest measure of the amount of space functionally occupied by an atom or residue in a protein. Furthermore, use of these normalized volumes allows one to compare cavity volumes calculated by different methods with those obtained from MD simulations (see below).

Although the MC and CAST procedures successfully detected all 43 model mutant cavities, VOIDOO and MS failed to detect cavities in one (V104A of staphylococcal nuclease (SNase) (1STN)) and six cases (F34A of SNase (1STN), F8A of RNase A (1RVU), W59A of RNase T1 (1BV1), I55A of hen egg white lysozyme (19SL), F56A and F57A of human lysozyme (1LZ1)), respectively. The cavity volumes determined by CAST in a few cases differ by ~30 Å³ from those determined by our method. Fig. 4 shows stereo representations of the cavity detected by the MC procedure in one such case (barnase I88A mutant (1BRJ)). The empty space or cavity formed because of the replacement of Ile → Ala in the barnase mutant (Fig. 4a) is shown to be detected well by the space-filled grid points of the MC method (Fig. 4c). Fig. 4b shows that the cavity occupies the same physical space as the wild type Ile residue side-chain atoms. In the other cases where discrepancy existed, visual examination suggested that cavity volumes calculated by the MC procedure were more appropriate than those calculated by CAST.

The reason for the differences in calculated cavity volumes in the above mutants detected is not apparent. To examine which cavity calculation (MC or CAST) was more accurate, we carried out MD simulations on eight different solvated cavity-containing mutants of T4 lysozyme, RNase S, and barnase, which showed significant differences in normalized cavity volumes calculated by the two procedures. The excess Voronoi volumes associated with atoms surrounding cavities in these mutants were calculated using solvated MD structures of the wild type and mutant proteins extracted at different time points during the course of simulation. These excess volumes are taken to be realistic estimates of cavity volume in the mutants and were compared with volumes calculated using the two methods.

**MD Simulations of Cavity-containing Mutants**—Molecular dynamics simulations of solvated cavity-containing mutants of barnase (L14A, I88A), RNase S (F8A, M13A), and T4 lysozyme (F153A, I58A, I100A, V103A), as well as for the corresponding wild type proteins, were carried out to obtain structures with energetically correct positioning of water molecules around the proteins (Table III). Voronoi volume calculations were per-

---

**Fig. 2. Comparison of residue volumes computed using MD structures and the MC method.** The percentage of change in residue Voronoi volume determined in the MC method with respect to those averaged over the course of an MD simulation is shown as a function of residue depth. a, RNase S; b, barnase; c, T4 lysozyme.
Energetics of Cavity Formation in Proteins

We have therefore calculated cavity volumes for a larger set of mutants with large → small residue replacements (Table II). Mutants of RNase S (28, 36–39, 41), T4 lysozyme (24, 40), barnase (42, 43), human lysozyme (44), and Gene V protein (25, 26) were chosen for this study. Only mutants with 15% or less accessibility in the wild type protein were considered for the study. Volume calculations were performed on the wild type as well as on the mutant proteins. Δ[cavity volume], the difference between the cavity volumes in the mutant and wild type, is shown to correlate with the loss of stability ΔG_{U→F} (r^2 = 0.71, slope = 37.5 ± 2.9 kcal mol^{-1} Å^{-3}, intercept = 0.59 ± 0.12 kcal mol^{-1}) in the MC method. A poorer correlation is seen with cavity volumes calculated using CAST (r^2 = 0.52, slope = 36.1 ± 4.2 kcal mol^{-1} Å^{-3}, intercept = 0.49 ± 0.30 kcal mol^{-1}) (Fig. 6, a and b). A high linear correlation between MC and CAST cavity volumes (r^2 = 0.87, slope = 0.88, intercept = 10 Å^3) was seen, indicating that cavity volumes detected by both methods are close to each other except in a few cases. In cases where there were large differences, MD simulations were carried out as described above. The results of the MD simulations clearly showed that the volumes calculated by the MC method were more accurate. ΔG_{U→F} showed a much poorer correlation (r^2 = 0.3 and 0.44) with cavity volumes calculated using the MS and VOIDOO programs, respectively (data not shown). This is probably caused by the fact that cavities close to the surface are often not detected by either program.

The loss of stability (ΔΔG_{U→F}) resulting from such large to small substitutions was partitioned into two components: 1) resulting from the loss of packing at the site of mutation and 2) resulting from the change in the hydrophobic driving force as a result of the mutation. The latter term is quantitated in terms of the change in the free energy (ΔΔG_{ASA}) of transfer of the amino acid at the site of mutation from solvent to the protein interior. The first component is proportional to the Δ[cavity volume] and the second to the ΔΔASA. ΔΔASA is the difference between the mutant and the wild type residue ASA buried upon folding. To get an estimate of the relative contributions of these two factors, the data were fit to the following equation.

\[
\Delta \Delta G_{U\rightarrow F} = m_1 \cdot [\Delta \text{cavity volume}] + m_2 \cdot [\Delta \Delta ASA]
\]

The values of m1 and m2 were 26.1 ± 5.6 cal mol^{-1} Å^{-3} and 18.5 ± 5.4 cal mol^{-1} Å^{-2}, respectively, in the MC method. The incorporation of the hydrophobic effect term improves the agreement between observed and calculated values of ΔΔG_{U→F}. In Fig. 5, the root mean square deviation difference between the volumes obtained by MC and MD methods is −10 Å^3. In an attempt to see the effect of an error in cavity volume estimate of 10 Å^3 on the fitted parameters in Equation 6, the following two exercises were carried out. (a) Each cavity vol-
The change in stability upon mutation is shown as a function of cavity volumes calculated by the MC and CAST procedures. Cavities in 1RBH, 1RBI, 1D5D, 1RBF, 1RC, 1D5H, 1BSE, 235L, 236L, 240L, 243L, 1OUB, 1OUJ, 1YAN, and 1YAO were not detected by the MS program. Cavities in 1RBH, 1RBI, 1BSE, 1BRH, 1BRJ, 235L, 236L, 240L, 243L, 1OUB, 1OUJ, 1YAN, 1YAO, 1YAM, and 1VQ1 were not detected by VOIDOO.

### Table II

**Energetics of cavity formation in proteins**

| Protein Data | Mutant | $\Delta G^0$ | $\Delta \text{cavity}$ |
|--------------|--------|--------------|-----------------------|
| Bank identification code |                      | kcal/mol | Å$^3$ |                      |
| RNase S | F8A | 5.1 | 130 | 160 |
| 1D5H$^a$ | M13G | 5.0 | 85 | 99 |
| 1RBF | M13A | 4.3 | 76 | 72 |
| 1RC | F8M | 3.6 | 76 | 93 |
| 1D5D | M13L | 0.6 | 57 | 53 |
| 1RBH | M13V | 0.3 | 18 | 24 |
| Barnase | L14A | 4.3 | 64 | 49 |
| 1BRH | I88A | 4.0 | 68 | 38 |
| 1BRJ | I96A | 1.9 | 51 | 51 |
| 1BSE | L89V | 0.3 | 0 | 0 |
| T4 lysozyme | L99A, F153A | 8.3 | 210 | 193 |
| 1LS4 | L98A | 5.0 | 102 | 108 |
| 1LS9 | L133A | 3.6 | 73 | 86 |
| 1LS5 | F153A | 3.5 | 113 | 112 |
| 244L | I100A | 3.4 | 73 | 42 |
| 243L | I58A | 3.2 | 57 | 28 |
| 240L | I27A | 3.1 | 26 | 32 |
| 200L | L121A | 2.7 | 44 | 46 |
| 239L | I17A | 2.7 | 43 | 56 |
| 1L67 | L46A | 2.7 | 38 | 31 |
| 238L | V103A | 2.2 | 41 | 38 |
| 236L | F67A | 1.9 | 35 | 52 |
| 235L | V87A | 1.7 | 47 | 41 |
| Human lysozyme | V111A | 1.3 | 6 | 29 |
| 1OUD | V121A | 1.5 | 49 | 52 |
| 1YAQ | I56A | 1.2 | 6 | 24 |
| 1YAP | I59A | 1.1 | 32 | 52 |
| 1OUJ | V99A | 1.0 | 35 | 46 |
| 1YAM | I106V | 0.7 | 21 | 0 |
| 1YAO | I58V | 0.5 | 18 | 24 |
| 1YAN | I27A | 0.4 | 2 | 26 |
| 1OUB | V100A | 0.3 | 0 | 21 |
| Gene V protein$^b$ | 1VQ1 | I47V | 1.0 | 28 | 40 |
| 1VQA | V35A, I47L | 0.6 | 13 | 30 |

$^a$ The cavity volume of 1D5H (F8A of RNase S) is corrected from the graphics as Phe-120 moves to fill up the space created by removal of Phe-8. The value of the new cavity generated due to the movement of Phe-120 is added to obtain the total cavity volume.

$^b$ The $\Delta G^0_{\text{U-F}}$ and cavity volumes of Gene V protein are shown per monomer.

A methylene group has average values of 27 Å$^3$ and 22 Å$^2$ for Voronoi volume and ASA, respectively. Thus, deletion of a single methylene group in the absence of any structural rearrangement would result in a decrease in stability of 0.7 and 0.41 kcal mol$^{-1}$ because of loss of packing and hydrophobicity, respectively. Hence, the magnitude of the contribution made by a single methylene group is 1.1 kcal mol$^{-1}$. This is in excellent agreement with the average value of 1.1 kcal mol$^{-1}$ per methylene group reported from cavity-filling mutations (45) but is slightly lower than the average value of 1.3 kcal mol$^{-1}$ reported earlier for another data set of cavity-creating mutations (46). In the present data set, there are mutants with $\Delta G^0_{\text{U-F}}$ either higher or lower than the corresponding average values, e.g., L99A of T4 lysozyme ($\Delta G^0_{\text{U-F}} = 5.0$ kcal mol$^{-1}$) and I17A and L46A ($\Delta G^0_{\text{U-F}} = 2.7$ kcal mol$^{-1}$). In such cases the cavity volumes are either larger or smaller than the corresponding expected cavity volume, i.e., difference between the side-chain Voronoi volume of wild type and the mutant residue (74 Å$^3$ in this case).

From the fitted values described above, the approximate maximal values associated with deletion of one (e.g., Ile → Val), two (e.g., Val → Ala) or three methylene (e.g., Ile → Ala, Leu → Ala) are 1.1, 2.2, and 3.3 kcal/mol, respectively. However, it should be remembered that the packing contribution obtained above is an average value and in general will depend on the actual packing density around the site of mutation. The close agreement between observed and calculated values of $\Delta G^0_{\text{U-F}}$ at several sites in multiple proteins suggests that the variation in packing density is not large, at least for the present set of mutational data.

There are several mutants of SNase such as L108A, I72A, V111A, L125A, and L103A, where the experimentally observed values of $\Delta G^0_{\text{U-F}}$ (5.8, 5.1, 4.7, 4.9, and 4.6 kcal mol$^{-1}$) are larger than expected from the present analysis (27). Cavity volume calculations of these mutants could not be carried because structures are not available. The stability changes for
these mutants were obtained using guanidine hydrochloride induced denaturation and assuming two-state unfolding (27). The m-values (an index of the cooperativity of folding) of these mutants are substantially less than that of the wild type. Subsequently Carra and Privalov (47) have shown that the SNase mutants of this nature, termed as m/\(H_{11002}\) mutants, thermally denature via a three-state mechanism rather than a two-state mechanism. It was shown for some of the mutants that \(\Delta G_0^{U \rightarrow P}\) determined from calorimetric studies (with the three-state assumption) was substantially smaller in magnitude than the \(\Delta G_0^{U \rightarrow P}\) previously determined from guanidine hydrochlorine denaturation studies (Table II of Carra and Privalov (Ref. 47)). Hence, consistent with the present work, the true \(\Delta G_0^{U \rightarrow P}\) change of the above SNase mutants is likely to be in the range observed for the proteins examined here.

Earlier studies by Eriksson et al. (24) showed that destabiliza-

![FIG. 4. A stereoplot of the barnase I88A structure (Protein Data Bank identification code 1BRJ) in the region around the cavity. Atoms within a sphere of radius of 7.0 Å from the centroid of the wild type Ile-88 are shown in the mutant structure. The van der Waals dot surface of these atoms around the cavity is shown to highlight the cavity that is formed. The CA and CB atoms of Ala-88 are shown in black. a, cavity is seen in the mutant protein. b, the side-chain atoms of Ile-88 (generated by superposing the wild type structure on the mutant protein) are shown to fill the space, indicating that the wild type protein is devoid of a cavity at this site. c, space-filled grid points (MC method) are shown to fill the empty space of the cavity, indicating proper detection of cavity.](image)

### TABLE III

Comparison of cavity volumes from MD simulations with those calculated by the MC and CAST procedures

| Protein     | Mutation | Residue depth | MC volume | CAST volume | MD volume |
|-------------|----------|---------------|-----------|-------------|-----------|
| Barnase     | 1BRH     | L14A          | 6.8 Å     | 65 Å^3      | 49 Å^3    | 60 Å^3    |
|             | 1BRJ     | I88A          | 9.3 Å     | 68 Å^3      | 38 Å^3    | 80 Å^3    |
| RNase S     | 1D5H     | F8A^a         | 8.0 Å     | 40 Å^3      | 56 Å^3    | 42 Å^3    |
|             | 1RBC     | M13A          | 5.9 Å     | 76 Å^3      | 72 Å^3    | 59 Å^3    |
| T4 lysozyme | 1L85     | F153A         | 10.1 Å    | 113 Å^3     | 112 Å^3   | 118 Å^3   |
|             | 244L     | I100A         | 6.0 Å     | 73 Å^3      | 42 Å^3    | 78 Å^3    |
|             | 243L     | I58A          | 6.8 Å     | 57 Å^3      | 28 Å^3    | 71 Å^3    |
|             | 238L     | V103A         | 6.0 Å     | 41 Å^3      | 38 Å^3    | 54 Å^3    |

^a^ Refers only to the cavity in the region of Phe-8 and does not include the cavity near Phe-120.
tion of several T4 lysozyme Leu → Ala cavity-containing mutants was approximated by a constant term (−2.0 kcal mol⁻¹) plus a term that increases in proportion to the size of the cavity (24). Cavity volumes in this study were measured by the MS procedure. The constant term arising as a result of the loss of three -(CH₂)₃ groups in Leu → Ala mutants was shown to be approximately equal to the transfer free energy of leucine relative to alanine as determined from partitioning between aqueous phase and octanol. Hence, for a single -(CH₂)₃, this would correspond to ≈ 0.6 kcal mol⁻¹ and is slightly larger than our estimate of 0.41 kcal mol⁻¹. The estimate of the cavity size-dependent energy term by Eriksson et al. (24) was shown to range from 24 to 33 cal mol⁻¹ Å⁻³. Our estimate of the cavity dependent energy term of 26.1 cal mol⁻¹ Å⁻³ is in numerical agreement with this previous result, although it should be noted that the volumes measured in the MS procedure (before normalization) are ~50% larger than the Voronoi volumes obtained after normalization by the MC procedure (Table I). Our estimate of the strength of the hydrophobic driving force is similar to but somewhat lower than the value of 25 cal mol⁻¹ Å⁻² obtained in those studies, but is obtained using a much larger collection of mutants in five different proteins. Eisenhaber (48) reported a similar value of 18 cal mol⁻¹ Å⁻² for the specific free energy of solvation of hydrophobic surface regions of proteins, which was estimated from the area distribution of hydrophobic surface patches. Recently a series of MD simulations of aliphatic solutes in water (49) was used to obtain a microscopic estimate of the hydrophobic driving force of 24 cal mol⁻¹ Å⁻².

The present work shows that our previous attempt (28) to correlate values of ΔG°(U→P) with cavity volumes was unsuccessful because the methods of cavity detection used earlier failed to accurately detect and quantitate the cavity spaces. In a very recent study by Sundberg et al. (50), the decrease in affinity between hen egg white lysozyme and its antibody D1.3 because of substitution of an interface hotspot residue with smaller residue was shown to correlate with loss of apolar buried surface area caused by the mutation (50). The correlation between change in the binding free energy and the change in apolar buried surface area corresponded to 21 cal mol⁻¹ Å⁻² for the effective hydrophobicity at the mutation site residue 92 of D1.3. Our estimate of 18.5 cal mol⁻¹ Å⁻² for the hydrophobicity is very close to their estimate but quite different from the estimate of 47 cal mol⁻¹ Å⁻² proposed earlier (4, 21, 22). An earlier analysis that yielded this higher estimate was based on thermodynamic data from small molecule transfer studies (4). Although the analysis was carefully done, several implicit assumptions were involved, the most important being the applicability of such data to protein folding/stability studies. Although this analysis accounts very well for the maximal changes in stability that arise from small to large mutations, it is not able to satisfactorily account for the large amount of mutational data which have changes smaller than predicted by such an analysis. The analysis predicts that the loss of stability caused by removal of a -(CH₂)₃ group is approximately 2 kcal mol⁻¹, i.e. the loss of stability is approximately 6 kcal mol⁻¹ for Leu → Ala mutations (Table II of Nicholls et al. (Ref. 4)). A value of 6 kcal mol⁻¹ is an extreme value for an Leu → Ala mutation and is certainly greater than the average of approx-

---

**Fig. 5.** The average increase in the Voronoi volume of atoms surrounding cavities calculated from the MD simulations is plotted against cavity volumes determined by MC (a) and CAST (b). The straight lines show best fit of linear functions to the data.

**Fig. 6.** The dependence of ΔG°(U→P) on cavity volume created because of large → small substitutions. Circles represent the cavity volume calculated from the MC method (a) and CAST (b). The line shows a best fit of a linear function to the data. c, correlation between ΔG°(U→P) and ΔG°(U→P)MC. The line has the equation γ = x and is used to visualize the deviation between the observed and calculated values of ΔG°(U→P).
Cavities and the Hydrophobic Effect

approximately 3.2 kcal/mol (in the examples quoted by Nicholls et al. (Ref. 4)) or 3.6 kcal/mol for the examples examined here.

Cavities play a crucial role in protein function, dynamics, stability, and ligand binding. Even in proteins of known three-dimensional structure, accurate identification and quantitation of cavities is not straightforward. This is because proteins are not solid objects and most atoms in a protein are connected to the surface by narrow channels. We have developed a simple, Voronoi-based procedure for calculating volumes of atoms as well as cavities in proteins. The method is able to detect cavities not detectable by other methods. Volumes calculated by the present (MC) procedure are in good agreement with accurate residue and cavity volumes calculated from computationally expensive molecular dynamics simulations.

The hydrophobic effect is widely asserted to be one of the most important driving forces in biological phenomena such as protein folding and macromolecular association. Most estimates of the hydrophobic driving force are obtained from small molecule transfer studies. It has been difficult to obtain direct quantitative estimates of the strength of the hydrophobic effect from thermodynamic studies of actual proteins because the overall free energy of folding, \( \Delta G^0_{U \rightarrow F} \), has many contributions. The present cavity detection procedure was used to analyze structural and thermodynamic data for a large data set of cavity-containing mutants. The observed changes in \( \Delta G^0_{U \rightarrow F} \) upon mutation correlated better with cavity volumes calculated using the present procedure than with other procedures. The analysis yielded relative estimates of changes in packing and the hydrophobic driving force to the observed stability change. A value of \( \text{~18.5 cal mol}^{-1} \text{Å}^{-2} \) was obtained for the hydrophobic driving force. This is at the lower end of estimates of the driving force. This is at the lower end of estimates of the driving force.

Acknowledgments—We thank Dr. Jie Liang and A. Binkowski for helping us obtain the CAST results. We thank B. S. Sanjeev and A. Madhumalar for useful discussions on MD simulations and Dr. N. V. Joshi for discussions on statistical analysis.

REFERENCES

1. Richards, F. M. (1974) J. Mol. Biol. 82, 1–14
2. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehius, M. H. (1985) Science 229, 834–838
3. Bueh, M. J., Lattman, E. E., and Rose, G. D. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 4195–4199
4. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296
5. Lee, C., Park, S. H., Lee, M. Y., and Yu, M. H. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7727–7731
6. Kleywegt, G. J., and Jones, T. A. (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 178–185
7. Connolly, M. L. (1983) Science 221, 709–713
8. Connolly, M. L. (1993) J. Mol. Graph. 11, 139–141
9. Liang, J., Edelbrunner, H., Fu, P., Sudhakar, P. V., and Subramanian, S. (1998) Proteins 33, 1–17
10. Liang, J., Edelbrunner, H., and Woodward, C. P. (1998) Protein Sci. 7, 1884–1897
11. Liang, J., Edelbrunner, H., Fu, P., Sudhakar, P. V., and Subramanian, S. (1998) Proteins 33, 18–29
12. Chakravarty, S., and Varadarajan, R. (1999) Struct. Fold. Des. 7, 723–732
13. Voronoi, G. (1907) J. Reine Angewandte Mathematik 133, 97–178
14. Gerstein, M., and Chothia, C. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 10167–10172
15. Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999) J. Mol. Biol. 290, 253–266
16. Gerstein, M., Tsai, J., and Levitt, M. (1995) J. Mol. Biol. 249, 955–966
17. Dill, K. A. (1990) Biochemistry 29, 7133–7155
18. Chothia, C. (1974) Nature 248, 338–339
19. Richards, F. M. (1977) Annu. Rev. Biophys. Bioeng. 6, 151–176
20. Tanford, C. (1980) The Hydrophobic Effect, 2nd Ed., John Wiley & Sons, New York
21. Sharp, K. A., Nicholls, A., Fine, R. F., and Honig, B. (1991) Science 252, 106–109
22. Sharp, K. A., Nicholls, A., Friedman, R., and Honig, B. (1991) Biochemistry 30, 9696–9697
23. Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) Biochemistry 28, 4914–4922
24. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Science 255, 176–183
25. Sandberg, W. S., and Terwilliger, T. C. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 1706–1710
26. Sandberg, W. S., and Terwilliger, T. C. (1991) Trends Biotechnol. 9, 59–63
27. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) Biochemistry 29, 8033–8041
28. Ratnaparkhi, G. S., and Varadarajan, R. (2000) Biochemistry 39, 12365–12374
29. Makhatsadze, G. I., and Privol'es, P. L. (1995) Adv. Protein Chem. 47, 307–425
30. Harpaz, Y., Gerstein, M., and Chothia, C. (1994) Structure 2, 641–649
31. Kearsley, S. K. (1989) Acta Crystallogr. Sect. A 45, 208–210
32. Hubbard, S. J., and Argos, P. (1995) Protein Eng. 8, 1011–1015
33. Lee, B., and Richards, F. M. (1971) J. Mol. Biol. 55, 397–400
34. Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibl, G. L., Singh, U. C., Weiner, P., and Kollman, P. A. (2000) AMBER 5.0, University of California, San Francisco
35. Gellaty, B. J., and Finney, J. L. (1982) J. Mol. Graph. 1, 177–185
36. Connolly, M. L. (1993) J. Mol. Biol. 61, 305–322
37. Kim, E. H., Varadarajan, R., Wyckoff, H. W., and Richards, F. M. (1992) Biochemistry 31, 12304–12314
38. Varadarajan, R., Connolly, M. L., Sturtevant, J. M., and Richards, F. M. (1990) Biochemistry 29, 6108–6114
39. Xu, J., Baase, W. A., Baldwin, E., and Matthews, B. W. (1998) Protein Sci. 7, 158–177
40. Thomson, J., Ratnaparkhi, G. S., Varadarajan, R., Sturtevant, J. M., and Richards, F. M. (1994) Biochemistry 33, 8587–8593
41. Buckle, A. M., Henrick, K., and Fersht, A. R. (1993) J. Mol. Biol. 234, 847–860
42. Buckle, A. M., Cramer, P., and Fersht, A. R. (1996) Biochemistry 35, 4298–4305
43. Takano, K., Ogashara, K., Kaneda, H., Yamagata, Y., Fujii, S., Kanaya, E., Kikuchi, M., Oobatake, M., and Yutani, K. (1995) J. Mol. Biol. 254, 62–76
44. Akasaka, A., Haruki, M., Oobatake, M., and Kanaya, S. (1997) J. Biol. Chem. 272, 18686–18693
45. Pace, C. N. (1992) J. Mol. Biol. 226, 29–35
46. Carr, J. H., and Privol'es, P. L. (1995) Biochemistry 34, 2034–2041
47. Eisenhaber, F. (1996) Protein Sci. 5, 1676–1686
48. Raschke, T. M., Tsai, J., and Levitt, M. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 5965–5969
49. Sundberg, E. J., Urrutia, M., Braden, B. C., Isen, J., Tsuchiya, D., Fields, B. A., Malchiodi, E. L., Tormo, J., Schwarz, F. P., and Mariuzzo, R. A. (2000) Biochemistry 39, 15375–15387