Biochemistry

 Volume 26, Number 22 November 3, 1987

Perspectives in Biochemistry

Genetic and Structural Analysis of the Protein Stability Problem[†,‡]

Brian W. Matthews
Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403
Received June 12, 1987; Revised Manuscript Received July 28, 1987

The Problem

It has long been known that folded proteins are only marginally stable. There is a delicate balance between stabilizing interactions, principally due to the hydrophobic effect, and destabilization due to the loss of conformational entropy of the folded protein. The net difference between the free energies of the folded and unfolded forms is small, e.g., 5–20 kcal/mol. The understanding of protein stability is therefore complicated not only by the size and complexity of macromolecules but also by the necessity to account for very small energy changes in both the folded and the unfolded forms. A change in energy of a few kilocalories per mole in *either* the folded *or* the unfolded form of a protein can substantially alter its stability [see, e.g., Creighton (1984) for an excellent overview].

This short review describes the use of mutant proteins in studying the problem of protein stability. The emphasis is on work from the author's laboratory on the lysozyme from bacteriophage T4, studies made possible thanks to the insight and imagination of the late George Streisinger.

Random Mutagenesis

A traditional approach to determining the roles of individual amino acids in protein stability has been through the analysis of proteins derived from random mutagenesis. For brevity, "mutant protein" is used to describe a protein that is the product of a mutated gene. The best known example of random mutagenesis is provided by the mutant human hemoglobins. Randomly created mutants of a number of other proteins have been generated by genetic screens that allow the identification of variants with altered properties such as enzymatic activity, thermal stability, or DNA affinity [e.g., see Streisinger et al. (1961)]. Mutant proteins identified by these screening procedures are especially useful in highlighting individual residues that are important for function or stability. There is no preconceived bias, as is often the case with "site-directed" mutagenesis.

Analyses of mutant proteins are, of course, most informative in the context of known three-dimensional structures. Such studies suggest that there is no simple pattern in the chemical identity and location of amino acid substitutions that alter protein stability (Grütter et al., 1979, 1983; Perutz, 1980; Alber et al., 1986, 1987a; Pakula et al., 1986; Shortle & Lin, 1985). Observed amino acid substitutions result in changes in hydrophobicity, hydrophilicity, charge, and bulk. Both internal and external sites on the protein are subject to destabilizing changes. The consequences of a given mutation depend on the nature of the amino acid substitution and the environment in which it occurs. These results strongly suggest that many different types of noncovalent interactions including ion pairs, hydrogen bonds, van der Waals contacts, and hydrophobic contacts can all make quantitatively comparable contributions to the stability of a protein [e.g., see Brandts (1967) and Matthews et al. (1974)].

Notwithstanding the above variability, destabilizing mutations may have one feature in common. In the case of T4 lysozyme, known temperature-sensitive mutations occur at sites that are of low mobility and low solvent accessibility in the folded protein (Alber et al., 1987a). Mutations that substantially destabilize the protein apparently do so by eliminating, or interfering with, specific favorable interactions that occur in the folded wild-type protein. Because of the difference in entropy, well-defined intramolecular interactions can have substantially lower free energies than corresponding intermolecular interactions. Creighton (1983) has discussed this effect in terms of the concept of "local concentration". As an example, a well-defined hydrogen bond between thermolysin and a bound inhibitor provides 4.0 kcal/mol in binding energy, even though the dissociated enzyme and inhibitor can make "equivalent" hydrogen bonds to the solvent (Bartlett & Marlowe, 1987; Tronrud et al., 1987).

Another intriguing aspect of the observed locations of ts mutations in phage lysozyme concerns possible implications for the unfolded state. Suppose that a class of mutations caused temperature sensitivity by increasing the stability of the unfolded protein (or by favoring the formation of some nonnative form). Amino acid substitutions of this sort could,

[†]This work was supported in part by grants from the National Institutes of Health and the National Science Foundation.
[‡]Dedicated to the memory of the late George Streisinger.

0006-2960/87/0426-6885$01.50/0

presumably, occur anywhere in the sequence, including the flexible parts of the folded protein. Because no destabilizing mutations have been observed in the mobile, surface-exposed, residues of T4 lysozyme, such mutations must be rare or nonexistent. There are mutants of T4 lysozyme that are thought to change the free energy of the unfolded state by 1–2 kcal/mol, principally through entropic effects (see below), but there is no evidence for mutants that cause *large* changes in the free energy of the unfolded protein. Obviously the unfolded state cannot be ignored, but the prediction of the behavior of mutant proteins is greatly simplified if their energetics are dominated by their respective folded structures. An alternative view has, however, been advocated by Shortle and co-workers. On the basis of the solvent denaturation of mutants of staphylococcal nuclease Shortle and Meeker (1986) concluded that single amino acid substitutions act by altering interactions that occur in the *denatured* state. This inference is predicated on the assumption that the main consequence of point mutations is to alter the surface area exposed to solvent in the denatured state of the mutant protein.

Consistent with the idea that the thermodynamic stabilization of a protein is a global property, the free energy contributions to thermostability of independent mutations are usually additive, at least qualitatively (Matsumura et al., 1986; Baase et al., 1986; Hecht et al., 1986).

Crystallographic analyses of mutant proteins show that destabilizing (and nondestabilizing) amino acid substitutions usually cause small, localized alterations in the three-dimensional conformation of the protein (Grütter et al., 1979, 1983; Fermi & Perutz, 1981; Alber et al., 1986; Katz & Kossiakoff, 1986; Villafranca et al., 1987). Typically, a few side-chain atoms at or close to the site of substitution may move up to 1 Å. Associated adjustments in the backbone atoms are often 0.5 Å or less. Some mutations cause the displacement or addition of bound water molecules [e.g., Grütter et al. (1987) and Gray and Matthews (1987)]. Clearly, the stability of a protein can be modified without large-scale disruption of its tertiary structure.

One potential criticism of this conclusion is that the mutant proteins subject to crystallographic analysis are selected (or induced) by the crystallization process to have structures similar to those of their progenitors. In the case of T4 lysozyme, about 70 mutant proteins have been purified to date and, of these, about 80% have crystallized isomorphously with the wild-type protein (Alber & Matthews, 1987, and unpublished results). At least in this case a sizable fraction of all point mutants can be analyzed crystallographically. Of the remaining 20%, some fail to crystallize and others give different crystal forms. There are many examples in the literature suggesting that the structures of proteins are not significantly altered by crystallization [e.g., see Matthews (1976)]. In the case of T4 lysozyme, isomorphous crystals have been obtained from several mutants with conformational changes that propagate 10–15 Å through the structure [e.g., Alber et al. (1986) and Gray and Matthews (1987)]. Nevertheless, it is important to determine the structures of mutant proteins that give new crystal forms since any examples of especially large structural change will presumably be in this category.

Site-Directed Mutagenesis

It goes without saying that site-directed mutagenesis provides an extremely powerful new tool with which to address the protein stability problem. As an example I take the liberty of quoting results obtained by Alber et al. (1987b) in the author's laboratory.

One of the temperature-sensitive lysozymes of phage T4 generated by random mutagenesis has the amino acid substitution Thr-157 → Ile (Grütter et al., 1987). In wild-type lysozyme the γ-hydroxyl of the threonine side chain participates in a network of hydrogen bonds that is disrupted in the mutant protein by the substitution of isoleucine. Loss of this hydrogen-bond network appears to be the main reason why the mutant lysozyme is less stable than wild type. Other changes between the two structures could, however, also contribute to instability (Grütter et al., 1987). To determine, in detail, the ways in which Thr-157 contributes to the stability of T4 lysozyme, 13 different amino acids have been substituted at this site. The structures of these modified lysozymes have been determined and their stabilities measured (Alber et al., 1986, 1987b; Alber & Matthews, 1987). The results show clearly that the main way in which Thr-157 contributes to stability is through its hydrogen-bonding interactions. An interesting situation occurs when glycine is substituted at position 157. The lack of a side chain allows a water molecule to bind at the site previously occupied by the γ-hydroxyl of the threonine and to restore the hydrogen-bond network, giving a protein whose stability is close to that of wild type. Solvent molecules bound on the surface of a protein obviously can contribute to protein stability and should not be ignored in the evaluation of protein energetics.

In the case of the 13 different substitutions that have been made at position 157 in T4 lysozyme, no mutant is more stable than wild type. This is, perhaps, not surprising. However, it was striking to find that no engineered mutant protein was of lower stability than Ile-157, the variant that was obtained as a temperature-sensitive mutant after random chemical mutagenesis. This tends to suggest that mutants selected by genetic screens as "temperature sensitive" are atypical in the sense that they are unusually destabilizing. In contrast, of all amino acid substitutions that are theoretically possible, the majority may have a relatively small effect on the stability of a protein. This conclusion is also supported by multiple substitutions at other sites in T4 lysozyme (Alber et al., 1987c, and unpublished results). According to this view, proteins are seen as tolerant of change and, within reason, relatively resistant to destabilization by amino acid replacements. It would be no surprise if protein structures had evolved the capacity to accommodate amino acid substitutions that occur during evolution (Alber et al., 1986).

Protein Stabilization

In addition to its commercial relevance, the ability to engineer proteins of enhanced thermostability is one of the prime tests of our understanding of protein energetics.

Most of the effort to increase the thermal stability of proteins by site-directed mutagenesis has focused on the introduction of nonnative disulfide cross-links. This has proven successful in some instances but not others (Perry & Wetzel, 1984; Sauer et al., 1986; Pantoliano et al., 1987; Villafranca et al., 1983, 1987; Wells & Powers, 1986; Bryan et al., 1985). Analysis of the results is complicated because of the different ways in which stability is measured. An engineered disulfide bond in dihydrofolate reductase increases stability with respect to unfolding with guanidine hydrochloride but decreases stability toward thermal denaturation (Villafranca et al., 1987). Disulfide linkages in known protein structures are observed to have restricted geometries (Richardson, 1981; Pabo & Suchanek, 1986; M. Levitt, personal communication). It turns out that relatively few pairs of amino acids in a given protein have the requisite separation and relative alignment to accept a connecting disulfide linkage without requiring adjustment of the host structure. In practice, pairs of cysteines introduced

at "nonideal" sites readily form –S–S– linkages, but the entropic stabilization gained by the formation of the cross-link may be offset by the energy required to distort the protein structure.

A recent approach to protein stabilization utilizes amino acid substitutions that decrease the configurational entropy of unfolding of the polypeptide backbone (Matthews et al., 1987). A glycine has greater backbone configurational entropy than an alanine and so requires more free energy to transfer from the unfolded to the folded state. By the same token, a proline has less backbone configurational entropy than any other amino acid and so requires less free energy to fold. By judiciously replacing glycines or inserting prolines at sites that do not interfere with the three-dimensional structure of a protein, it should be possible to decrease the entropy of unfolding of a protein and thereby increase its thermostability. This has been tested in T4 lysozyme by two different substitutions, Gly-77 → Ala and Ala-82 → Pro. Both substitutions increase the stability of the protein toward reversible and irreversible thermal denaturation at physiological pH (Matthews et al., 1987). The combination of a number of such amino acid substitutions, each of which is expected to contribute about 1 kcal/mol to the free energy of folding, should permit cumulative enhancement of protein stability.

Amino acid replacements of the form Gly → Ala have also been shown to increase the thermal stability of λ repressor (Hecht et al., 1986) and the neutral protease from *Bacillus stearothermophilus* (Imanaka et al., 1986). In these instances the enhanced thermostability was attributed to the replacement within an α-helix of a poor helix-forming residue (glycine) with a good helix former (alanine). The concept of entropic stabilization provides another rationalization for this effect.

A different type of helix stabilization is possible from favorable interactions with the helix dipole (Blagdon & Goodman, 1975; Hol et al., 1978). Mitchison and Baldwin (1986) have shown that peptide analogues of the S-peptide of ribonuclease are successively made more stable as the charge on the amino-terminal residue is changed from +2 to −1. Reconstitution of these semisynthetic S-peptide analogues into ribonuclease S also enhances the stability of the protein as a whole. This encouraging result clearly suggests that appropriate site-directed substitutions of charged residues at the ends of helices may enhance protein stability. Other strategies for stabilization of proteins of known structure include the addition of new interactions in rigid parts of the structure (Alber et al., 1987a), elimination (or satisfaction) of any unsatisfied hydrogen-bonding groups within the protein, and the elimination of any interior cavities.

In several cases genetic screens to identify proteins with enhanced thermostability have been developed (Alber & Wozniak, 1985; Matsumura & Aiba, 1985; Shortle & Lin, 1985; Liao et al., 1986; Bryan et al., 1986).

Summary

Substantial progress has been made recently in understanding the nature of protein stability. Recent evidence suggests that protein structures are catholic in the sense that they can accommodate different amino acids at many different sites. If the wild-type protein is relatively stable, such substitutions allow the formation of folded functional proteins with only modest changes in stability. This is particularly encouraging for those interested in the application of site-directed mutagenesis. Large changes in the stability of a protein due to the substitution of one amino acid by another are seen to be manifested primarily at the level of the folded rather than the unfolded structure. Loss of stability in structurally characterized ts mutants can be explained in terms of the loss of favorable intramolecular interactions and/or the introduction of unfavorable interactions that cannot be accommodated in the folded protein. Assumptions of major changes in the energy of the unfolded state are not required. This is encouraging for those interested in the calculation of the energetics of mutant proteins, although it clearly is essential to include the contributions of solvent and entropy in such calculations. There has been some success in increasing the thermostability of proteins by site-directed mutagenesis. New strategies for protein stabilization are being tested, and substantial progress in this area is to be expected in the near future.

Acknowledgments

The results with T4 lysozyme described here were obtained by T. Alber, W. J. Becktel, J. Bell, S. Dao-pin, T. M. Gray, M. G. Grütter, H. Nicholson, L. H. Weaver, K. Wilson, and J. A. Wozniak. In addition to these colleagues I thank Drs. T. Alber, W. J. Becktel, F. W. Dahlquist, M. Matsumura, J. A. Schellman, P. H. von Hippel, and the late George Streisinger for advice and helpful discussions on the nature of protein stability and for comments on the manuscript.

Registry No. Lysozyme, 9001-63-2.

References

Alber, T., & Wozniak, J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 747–750.

Alber, T., & Matthews, B. W. (1987) *Methods Enzymol.* (in press).

Alber, T., Grütter, M. G., Gray, T. M., Wozniak, J., Weaver, L. H., Chen, B.-L., Baker, E. N., & Matthews, B. W. (1986) in *UCLA Symp. Mol. Cell. Biol., New Ser. 39*, 307–318.

Alber, T. A., Dao-pin, S., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987a) *Biochemistry 26*, 3754–3758.

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P., & Matthews, B. W. (1987b) *Nature (London)* (in press).

Alber, T., Bell, J. A., Dao-pin, S., Nicholson, H., Wozniak, J. A., Cook, S. P., & Matthews, B. W. (1987c) *Science (Washington, D.C.)* (submitted for publication).

Baase, W. A., Muchmore, D. C., & Becktel, W. J. (1986) *Biophys. J. 49*, 109a.

Bartlett, P. A., & Marlowe, C. K. (1987) *Science (Washington, D.C.) 235*, 569–571.

Blagdon, D. E., & Goodman, M. (1975) *Biopolymers 14*, 241–245.

Brandts, J. F. (1967) in *Thermobiology* (Rose, A. H., Ed.) pp 25–72, Academic, New York.

Bryan, P., Rollence, M., Pantoliano, M., Quill, S., Wood, J., Perna, L., Matthew, J., Hsiao, H., & Poulos, T. (1985) *J. Cell. Biochem., Suppl. 9B*, 92.

Creighton, T. E. (1983) *Biopolymers 22*, 49–58.

Creighton, T. E. (1984) *Proteins, Structures and Molecular Properties*, W. H. Freeman, New York.

Fermi, G., & Perutz, M. F. (1981) *Hemoglobin and Myoglobin*, Clarendon, Oxford.

Gray, T. M., & Matthews, B. W. (1987) *J. Biol. Chem.* (in press).

Grütter, M. G., Hawkes, R. B., & Matthews, B. W. (1979) *Nature (London) 277*, 667–669.

Grütter, M. G., Weaver, L. H., Gray, T. M., & Matthews, B. W. (1983) in *Bacteriophage T4* (Mathews, C. K., Kutter, E. M., Mosig, G., & Berget, P. M., Eds.) pp 356–360,

American Society for Microbiology, Washington, DC.
Grütter, M. G., Gray, T. M., Weaver, L. H., Alber, T., Wilson, K., & Matthews, B. W. (1987) *J. Mol. Biol.* (in press).
Hecht, M. H., Sturtevant, J. M., & Sauer, R. T. (1986) *Proteins: Struct. Funct. Genet. 1*, 43–46.
Hol, W. G. J., van Duijnen, P. T., & Berendsen, H. J. C. (1978) *Nature (London) 273*, 443–446.
Imanaka, T., Shibazaki, M., & Takagi, M. (1986) *Nature (London) 324*, 695–697.
Katz, B. A., & Kossiakoff, A. (1986) *J. Biol. Chem. 261*, 15480–15485.
Liao, H., McKenzie, T., & Hageman, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 576–580.
Matsumura, M., & Aiba, S. (1985) *J. Biol. Chem. 260*, 15298–15303.
Matsumura, M., Yasumura, S., & Aiba, S. (1986) *Nature (London) 323*, 356–358.
Matthews, B. W. (1976) *Annu. Rev. Phys. Chem. 27*, 493–523.
Matthews, B. W., Weaver, L. H., & Kester, W. R. (1974) *J. Biol. Chem. 249*, 8030–8044.
Matthews, B. W., Nicholson, H., & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Mitchison, C., Baldwin, R. L. (1986) *Proteins: Struct. Funct. Genet. 1*, 23–33.
Pabo, C. O., & Suchanek, E. G. (1986) *Biochemistry 25*, 5987–5991.
Pakula, A. A., Young, V. B., & Sauer, R. T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8829–8833.
Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F., & Poulos, T. L. (1987) *Biochemistry 26*, 2077–2082.
Perry, L. J., & Wetzel, R. (1984) *Science (Washington, D.C.) 226*, 555–557.
Perutz, M. F. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 317–322, Elsevier/North-Holland Biomedical, Amsterdam.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Sauer, R. T., Hehir, K., Stearman, R. S., Weiss, M. A., Jeitler-Nilsson, A., Suchanek, E. G., & Pabo, C. O. (1986) *Biochemistry 25*, 5992–5998.
Shortle, D., & Lin, B. (1985) *Genetics 110*, 539–555.
Streisinger, G., Mukai, F., Dreyer, W. J., Miller, B., & Horiuchi, S. (1961) *Cold Spring Harbor Symp. Quant. Biol. 26*, 25–30.
Tronrud, D. E., Holden, H. M., & Matthews, B. W. (1987) *Science (Washington, D.C.) 235*, 571–574.
Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N., & Kraut, J. (1983) *Science (Washington, D.C.) 222*, 782–788.
Villafranca, J. E., Howell, E. E., Oatley, S. J., Xuong, N.-H., & Kraut, J. (1987) *Biochemistry 26*, 2182–2189.
Wells, J. A., & Powers, D. B. (1986) *J. Biol. Chem. 261*, 6564–6570.

The Two Cultures: Chemistry and Biology[1]

Arthur Kornberg

Department of Biochemistry, Stanford University, Stanford, California 94305
Received July 14, 1987

Much of life can be understood in rational terms if expressed in the language of chemistry. It is an international language, a language for all of time, and a language that explains where we came from, what we are, and where the physical world will allow us to go. Chemical language has great esthetic beauty and links the physical sciences to the biological sciences. Unfortunately, the full use of this language to understand life processes is hindered by a gulf that separates chemistry from biology. This gulf is not nearly as wide as the one between the humanities and sciences on which C. P. Snow focused attention. Yet, chemistry and biology are two distinctive cultures and the rift between them is serious, generally unappreciated, and counterproductive.

The historical roots of chemistry and biology are intertwined in many places, and the conflicts between the chemistry and biology cultures go far back as well. A particularly fascinating example is the attempt to understand the fermentation of sugar to alcohol by the yeast cell. It had been known for over 6000 years that the insipid juice of the crushed grape can be transformed in a few days into an intoxicating, tasteful wine. But the nature of the fermentation remained a total mystery until nearly the 19th century, when it became known that the compound fermented by grape juice was sucrose and that the principal products were alcohol and carbon dioxide.

Alcohol fermentation could have been solved by any of the great chemists of the early 19th century. Berzelius of Sweden or Liebig and Wohler of Germany might have succeeded except for their neglect of biology and their failure to recognize the central role of yeast in fermentation. Pasteur, an equally great chemist, did appreciate the role of yeast, but his exaggerated immersion in biology caused him to neglect his chemical roots and miss discovering that enzymes are the vital force of fermentation. To what extent the yeast cell and its chemistry are responsible for making wine and brewing beer became the subject of one of the most protracted and vitriolic polemics in science.

From the 17th century on, since Antony von Leeuwenhoek first saw yeast cells in fermentation sediments, many believed these globules to be the driving force of the process. Against this view, the chemists, despite prescient ideas about catalysis of chemical reactions, were contemptuous of attributing the chemistry of fermentation to yeast cells. How tragic that these outstanding scientists who correctly focused on the chemical

[1] Adapted from a lecture at a meeting of the American Association for the Advancement of Science in Chicago on Feb 16, 1987.

0006-2960/87/0426-6888$01.50/0