# Additivity Principles in Biochemistry\*

## Ken A. Dill

From the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-1204

We cannot yet reliably fold proteins or RNA molecules by computer, predict ligand binding affinities, compute conformational transitions, or use the sequence information in the Human Genome very effectively to understand biomolecular function and disease. Why not? Perhaps some of our models in computational biology are based on flawed assumptions. Thermodynamic additivity principles are the foundations of chemistry, but few additivity principles have yet been found successful in biochemistry.

#### Driving Forces in Biochemistry: the Language of Free Energies

According to the principles of thermodynamics, to predict how molecules act we must account for the free energies. Free energies are expressed in the language of van der Waals interactions, hydrogen bonding, ion pairing, solvation and hydrophobic interactions, and entropies due to translations, vibrations, rotations, and configurations. Free energies or entropies of protein or RNA folding, mutations, enzyme kinetics, or ligand binding are often modeled as sums, based on *group additivities*,

$$\begin{split} \Delta G &= \Delta G_{\text{amino acid } 1} + \Delta G_{\text{amino acid } 2} + ... \Delta G_{\text{amino acid } n} \text{ or} \\ \Delta G &= \Delta G_{\text{CH}_2} + \Delta G_{\text{OH}} + \Delta G_{\text{aromatics}} + ... \Delta G_{\text{other substituents}} \text{ or} \quad (\text{Eq. 1}) \\ \Delta G &= \Delta G_{\text{backbone}} + \Delta G_{\text{side chain}} \end{split}$$

or free energy component additivities,

 $\Delta G = \Delta G_{\rm nonpolar \; solvation} + \Delta G_{\rm hydrogen \; bonding} +$ 

 $\Delta G_{
m PDB-derived statistical potentials} + \Delta G_{
m polar burial or electrostatic effects} +$ 

 $\Delta G_{
m van \ der \ Waals \ or \ packing \ effects} + \Delta G_{\Phi\Psi} \ -$ 

 $T\Delta S_{\text{conformational}} - T\Delta S_{\text{side chain rotations}} + \text{etc.}$  (Eq. 2)

or entropy component additivities.

$$\Delta S = \Delta S_{\text{translations}} + \Delta S_{\text{rotations}} + \Delta S_{\text{vibrations}} + \text{etc.}$$
 (Eq. 3)

Thermodynamic additivity is the principle that if two components, A and B, contribute independently to some process, then the total change in free energy (or enthalpy or entropy) is the sum of components,  $\Delta G = \Delta G_A + \Delta G_B$ . However, additivity only applies if components A and B are independent.

We call these sums "thermodynamic additivity models." Are there "true" independent free energies of a hydrogen bond, a salt bridge, or a hydrophobic contact that we could we add together to compute binding or folding free energies? Are enzyme rate accelerations sums of translational, rotational, and vibrational free energy terms? If a peptide, say tetraalanine, binds a protein, does the free energy equal four times that of the binding of alanine? Is protein folding the sum of oil-towater-like transfers of each amino acid? Can we add surface area-based solvation terms to molecular dynamics force fields? Are the conformational entropies of biopolymers simple sums of the monomer entropies or of backbone plus side chain entropies?

Modelers using expressions like Equations 1–3 must pay attention to decisions about the interactions and entropies. What is the right balance of interactions in a model? What is the relative importance of a hydrogen bond and a hydrophobic contact? How should we estimate packing and side chain energies and entropies? Is the protein interior like a hydrocarbon liquid, diketopiperazine crystal, or something else? What mathematical forms should we use for the interactions? Clearly, flawed answers to such questions can be sources of errors in models.

But perhaps some of our models in computational biology are failing at a deeper level. The concept of additivity is a fundamental premise (1, 2) that is widely taken for granted. It may be that *all* choices of parameters in Equations 1–3 might fail. Perhaps the problem is additivity itself. Perhaps it is inappropriate to sum free energies no matter what the relative weights and no matter what choices we make for the individual terms in it.

## A Major Culprit? The "4th Law" of Thermodynamics: the Assumptions of Independence and Additivity

Without additivity, chemistry would have limited predictive power (3).<sup>1,2</sup> Additivity has been called the 4th law of thermodynamics (3). For example, if the heat of formation of covalent compounds were not equal to the sum of the bond enthalpies (if the heat of formation of carbon dioxide were not equal to twice the heat of a C–O bond) then chemical equilibria and kinetics would not be predictable from simpler reactions. Every chemical equilibrium would require its own separate measurement. We could not look up bond energies in tables and compute the energetics of ATP cycles, the breakdown of glucose, or other equilibria.

Thermodynamic additivity principles could be equally important in biochemistry (noncovalent processes). When we design drugs using quantitative structure-activity relationship substituent constants, when we use single-site mutagenesis as a basis for protein engineering, when we design folding and binding algorithms based on models of amino acid partitioning, or when we compute the melting behavior of DNA and RNA as sums of nearest neighbor interaction energies, we assume additivity relationships. The search for additivity principles is based on the hope that we will be able to make independent measurements or calculations for subcomponents of a system and add them together using equations such as 1–3 to predict the structures and properties of biomolecules.

In chemistry, the "state" called "a carbon dioxide molecule" and the "state" called "carbon and oxygen atoms" are sharply defined and distinguishable from each other, differing in stability by tens of kcal/mol, and not much dependent on temperature or solvent conditions. However, for biology, as for polymer science, the individual monomer interactions are much

<sup>\*</sup> These minireviews will be reprinted in the 1997 Minireview Compendium, which will be available in December, 1997. This is the first article of five in the "Protein Folding and Assembly Minireview Series."

<sup>&</sup>lt;sup>1</sup> M. Laskowski, personal communication.

<sup>&</sup>lt;sup>2</sup> J. Schellman, personal communication.

weaker, involving noncovalent interactions nearer to thermal energies (kT = 0.6 kcal/mol at physiological temperatures). In these cases, broad ensembles of microscopic conformations comprise macroscopic "states" (denatured states, molten globules, folding intermediates and transition states, some bound complexes). States are sometimes not so distinct as they are in covalent chemistry, and simple additivity may not apply.

#### How Good Does Additivity Need to Be?

Whether an additivity assumption is good or poor depends on what errors we can tolerate. Suppose we want an energy function to model protein folding. If an energy function has errors totaling 10 kcal/mol for a protein of 100 amino acids, it will be useless for predicting the native structure. Ten kcal/mol is about the difference in free energy between native and denatured conformations, so such functions will have no meaningful ability to discriminate among protein conformations. On the other hand, an energy function with an error of 1 kcal/mol, even though not perfect, would be useful. Since random errors grow with the square root of the number of monomers (5), a 100amino acid protein has about  $\sqrt{100} = 10$  times the error of one amino acid, so an adequate energy function must err by less than about 100 cal/mol per amino acid. (Only 10 cal/mol would be tolerable for systematic errors, which scale linearly with size.) This is a crude estimate, but it gives a rough target: if we can model contact, transfer, or binding interactions to better than 100 cal/mol for molecules about the size of amino acids or nucleotides, then our energy functions will be useful for models of folding or other large conformational changes.

## Group Additivities in Biochemistry

Group additivities account successfully for the partitioning of alkanes, alkenes, alkadienes, alcohols, and other homologous series of hydrocarbons from one medium to another (6, 7). The free energy of transfer depends linearly on the number of monomer units in the chain. Additivity predicts effects of double mutations on enzyme reaction rates (8), binding, and protein stability (9) (see Fig. 1) as sums of single mutations, at least when the mutation sites are spatially separated. Sometimes mutational effects on protein stability correlate well with oil/water partitioning (10). The stabilities ( $\Delta G$ ) and melting behavior ( $\Delta H$ ) of oligomers of DNA (11) or RNA (12) are well predicted as sums of free energies and enthalpies of nearest neighbor pairs of nucleotides. Additivity predicts well mutational changes in the binding of proteinases to their inhibitors.<sup>3</sup>

But these successes of additivity may owe as much to *uniformity of the neighboring environment*, as to additivity *per se*. The remarkable linearities in homologous series may depend on the constancy of next neighbors; for long chains, each added substituent *has the same neighbors* as the previous substituent. Often the monomers and dimers in homologous series do not fall on the same line as the longer chains. The free energy of transferring the tripeptide Gly-Gly-Gly from oil to water minus the free energy of transferring Gly-Gly should equal the free energy of Gly-Gly minus the free energy of glycine, since both differences represent the transfer of a single glycine. However, Nozaki and Tanford (14) showed that the former difference is 1270 cal/mol while the latter is 895 cal/mol. This discrepancy of 375 cal/mol is larger than the target error of 100 cal/mol.

Moreover, the rank ordering of partitioning amino acids into oil is different for different oil phases (15), and solute partition coefficients depend on the "ordering" in oil phases (16). The additive free energies of binding observed in protease inhibi-



FIG. 1. Additivity: double-site mutational changes are well predicted as sums of single-site changes. *Top*, stability; *bottom*, binding. From Sandberg and Terwilliger (9), on gene V protein of bacteriophage f1.

tors<sup>3</sup> apply to a particular site. What happens at one site may differ from what happens at another. It is fundamental to biology that biochemical environments (in complexes, inside proteins and RNAs, at binding sites) differ in their structures and energetics. It may be that "effective medium" models (like oils or solids), where additivity applies, will not adequately model complex biochemical environments.

Non-additivity corrections are common. Next neighbor models for RNA and DNA apply only to subclasses of conformations; non-additivity corrections are required for internal loops, bulge loops, hairpins, tertiary interactions, cruciforms, and non-B-form DNA (11, 12). Quantitative structure-activity relationship substituent methods require several corrections: geometric and flexing factors, chain branching, electronic factors, hydrogen bonding, and polyhalogenation factors (17–19). Nonadditivity corrections to free energies are sometimes called "cooperativity" (20, 21) or "conditional free energies" (22).

The limit of current experimental errors is probably around 200- 400 cal/mol (see the deviations from the straight line in Fig. 1, for example). If so, it implies a possible fundamental limitation on our ability to find useful thermodynamic additivity principles in biochemistry.

#### Energy Component Additivities in Biochemistry

Even more problematic than group additivity is energy component additivity (1, 18). Mark and van Gunsteren (1) have

 $<sup>^3</sup>$  M. A. Qasim, P. J. Ganz, C. W. Saunders, K. S. Bateman, N. G. James, and M. Laskowski, Jr., submitted for publication.



FIG. 2. Enthalpy/entropy compensation in six different mutations. *Top*, binding of modified S-peptides to S-protein to form ribonuclease S; *bottom*, unfolding of ribonuclease H1. From Sturtevant (25).

proven that adding entropies or free energy terms, as in Equation 2, is generally not justified, although sometimes the non-additivities may be small.<sup>4</sup> More broadly justified is the summing of energies or enthalpies:  $\Delta H_{\text{total}} = \Delta H_{\text{van der Waals}} + \Delta H_{\text{solvation}} + \Delta H_{\text{electrostatics}} + \Delta H_{\text{hydrogen bonding}}$ , etc., provided the terms describe independent forces (1). If we combine this with  $\Delta G = \Delta H - T\Delta S$ , an additivity relationship that is always justified by thermodynamics, then we might hope that expressions such as  $\Delta G = \Delta H_{\text{van der Waals}} + \Delta H_{\text{solvation}} + \Delta H_{\text{electrostatics}} + \Delta H_{\text{hydrogen bonding}} - T\Delta S$  are on a sounder footing than sums of free energy components.

How should we divide the total enthalpy into truly independent component terms? For example, what experiments will isolate enthalpies of solvation from hydrogen bonding? Making isosteric mutations does not mean "no change" in van der Waals interactions, because two different shapes having the same volume can pack a cavity differently.

If we cannot yet predict free energies, it is even harder to predict enthalpies because of enthalpy/entropy compensation (24); perturbations that increase the enthalpy can also increase the entropy, with little or no effect on the free energy. Compensation, which occurs broadly throughout biochemistry, means that the enthalpy is not independent of the entropy. Fig. 2, taken from a review by Sturtevant (25), shows compensation for some mutants in two different systems: the unfolding of ribonuclease  $H_1$  and the binding of modified S-peptides to S-protein to form ribonuclease S. The figure shows that a free energy change due to a mutation can be less than 1 kcal/mol, while the corresponding enthalpy change can be 10 kcal/mol.

In another example, Breslauer *et al.* (26) bound netropsin and distamycin A to two very similar DNA molecules: A, ATATAT... on one strand and TATATA on the other and

<sup>4</sup> Brady, G. P., and Sharp, K. A. (1995) J. Mol. Biol. 254, 77-85.

B, AAAAAA on one strand and TTTTTT on the other. The free energy of binding netropsin is nearly identical, -12.7 kcal/mol to molecule A and -12.2 kcal/mol to molecule B, and the dependence on NaCl is similar in the two cases. Nevertheless the driving forces are completely different; binding to molecule 1 is dominated by enthalpy ( $\Delta H = -11.2$  kcal/mol), while binding to molecule 2 is dominated by entropy ( $\Delta H = -2.2$ 

# Repairing Incorrect Additivity Assumptions: the Role of Statistical Mechanics in Biochemistry

energies may lead to better microscopic models.

kcal/mol). A similar result is found for distamycin. A better understanding of the enthalpy and entropy components of free

While thermodynamic models such as Equations 1–3 require additivity assumptions, statistical mechanical models do not. Statistical mechanics provides rigorous tools for relating molecular structures to thermodynamic quantities. Statistical mechanical theories give a good accounting for some cooperativities in biochemistry. 1) The Zimm-Bragg and related models (27, 28) describe helix-coil transitions. The tendency of a monomer to form a helical turn is not independent of the helical tendency of its neighbor. 2) Ligand binding can involve a cooperativity of binding at different sites (20, 21). While thermodynamic additivity models often rely on untested assumptions (about independence, additivity, averaged medium models of the environment, or ways to lump degrees of freedom together) statistical mechanical models need not be limited in this way. Statistical mechanical models and atomic simulations can aim to identify all the relevant degrees of freedom without bias and to weight them according to the Boltzmann distribution law.

## Entropies Are Important in Biology; They Are Often Not Additive

An important class of non-additivities pertains to the entropies of conformational change. Polymer conformational entropies are often large and seldom additive. When chains obey random-flight statistics, as in denatured states, or when interactions are dominated by local factors, as in helical peptides, chain entropies are sums of monomer entropies (29). However, when nonlocal contacts are involved, in molten globules or folded or compact states, they are not. One contact between the chain ends can globally restrict the options available to all the monomers (30).

What is the error due to non-additivities? It can be large. If z is the number of conformational isomers per monomer ( $z \approx 3-10$ ), the entropy (per monomer) of a random-flight chain is approximately  $S = R \ln z$ , where R is the gas constant. But the "correction," the excluded volume entropy, when a chain is constrained to be compact is estimated to be (31)  $\Delta S = R \ln e = R$ , a free energy error of  $RT \approx 600$  cal/mol. This exceeds the target of 100 cal/mol. For a 100-mer protein, this "non-additivity" is about 60 kcal/mol (32, 33), a large driving force.

Non-additivities may be the rule for biopolymer conformational entropies and free energies. Theory predicts that the free energy of folding is not a simple sum of water-to-oil transfers of the constituent amino acids since the denatured state (under native conditions) can harbor some buried residues (34). The conformational entropy of the denatured state is not independent of external conditions; it depends on solvent and temperature since strongly denatured chains are expanded (high entropy), while under native conditions denatured states are compact (low entropy). The conformational entropy of protein folding is predicted not to equal the sum of backbone plus side chain entropies, *i.e.*  $\Delta S \neq \Delta S_{\rm sc} + \Delta S_{\rm bb}$  (35). As the backbone decreases its radius of gyration, it "freezes out" side chain conformations. Different loops, bulges, hairpins, and pseudoknots in DNAs and RNAs or disulfide-bonded regions,

hinges, flaps, and interfaces in proteins are predicted to be not independent of each other (36), as is assumed in random-flight models (30). Experiments have not yet tested these predictions. In many cases, non-additivities in entropies and free energies could be measured by summing state functions around a thermodynamic cycle to determine their deviations from zero.

Statistical potentials are amino acid contact pairing frequencies, derived from data bases, and are used in protein folding algorithms. They assume pairing free energies are additive:  $\Delta G = \Delta G_{\text{Ala-Ala}} + \Delta G_{\text{Gly-Tyr}} + \Delta G_{\text{Pro-Leu}} + \text{ etc. } (37, 38).$  However, model tests show that pairing frequencies are not independent (39). For example, if the hydrophobic amino acids drive protein collapse, they also indirectly drive polar groups to protein surfaces, so polar group pairings are not independent of the nonpolar group pairings.

Chemistry texts express entropies as sums of translations, rotations, and vibrations for small molecules in the gas phase. Are such sums,  $\Delta S = \Delta S_{\text{translations}} + \Delta S_{\text{rotations}} + \Delta S_{\text{vibrations}} + \Delta S_{\text{conformations}} + \Delta S_{\text{solvation}}$ , also justified for liquids, solvation, biomolecule binding, and enzyme reaction kinetics (40)? Remarkably little evidence bears on this question.

Statistical mechanical theory shows that in oil phases and other polymer solutions, translations are coupled to polymer conformations (*i.e.*  $\Delta S \neq \Delta S_{\text{translation}} + \Delta S_{\text{conformation}}$ ) (31, 41). In liquid crystalline solutions, rotations are coupled to translations (42, 43). Free volume may be coupled to translations in some solutions (23). In general the shapes of solutes and solvents affect the interdependence of their degrees of freedom (13, 41), but models of ligand binding and enzyme mechanisms depend on separability assumptions (40). If the hydrophobic effect involves restricted water orientations around nonpolar solutes, then the water translational and rotational degrees of freedom are coupled. Such considerations have led Sharp et al. (4) to argue that the hydrophobic effect may involve much more free energy (45 cal/mol  ${\rm \AA}^2)$  than previously thought (25 cal/mol  ${\rm \AA}^2$ ) (7). Since the binding of proteases to their inhibitors can involve 600–1800 Å<sup>2</sup> of contact area (2), such uncertainties can amount to more than 10 kcal/mol.

#### **Conclusions**

A wide class of models in computational biology assumes thermodynamic additivity and independence (of energy types. of neighbor interactions, of conformational freedom, of monomer contact pairing frequencies, etc.). Biomolecules may achieve stability in the face of thermal uncertainty, as polymers do, by compounding many small interactions, but this summing trick works against modelers, since it compounds our errors. Weak interactions imply ensembles of states and possible non-additivities of entropies and free energies.

If thermodynamic additivity principles can be found having variances smaller than about 0.1 kcal/mol of monomer units. they may be as important to biochemistry as the great symmetry principles are to physics. At the present time, however, additivity principles appear to be few and limited in scope in biochemistry. Neighborhood and environment effects on additivities need to be better understood. To measure non-additivities, experimentalists could test the closure of thermodynamic cycles. Statistical mechanical models and molecular dynamics

simulations may contribute to more predictive theories in biochemistry.

Acknowledgments-I thank Sarina Bromberg, Hue Sun Chan, Michael Laskowski, John Schellman, and Tom Terwilliger for very helpful discussions and the National Institutes of Health and ONR for support.

#### REFERENCES

- 1. Mark, A. E., and van Gunsteren, W. F. (1994) J. Mol. Biol. 240, 167-176
- 2. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr., and James, M. N. G. (1995) Protein Sci. 4, 1985-1997
- 3. Benson, S. W. (1976) Thermochemical Kinetics: Methods for Estimation of Thermochemical Data and Rate Parameters, 2nd Ed., John Wiley & Sons, Inc., New York
- 4. Sharp, K. A., Nicholls, A., Fine, R. M., and Honig, B. (1991) Science 252, 106 - 109
- 5. Bryngelson, J. D. (1994) J. Chem. Phys. 100, 6038-6045
- Rytting, L. P., Huston, L. P., and Higuchi, T. (1978) J. Pharm. Sci. 67, 615-618 7. Tanford, C. (1980) The Hydrophobic Effect, 2nd Ed., John Wiley & Sons, Inc., New York
- 8. Wells, J. A. (1990) Biochemistry 29, 8509-8517
- 9. Sandberg, W. S., and Terwilliger, T. C. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 8367-8371
- 10. Ericksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Science 255, 178-183
- 11. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 3746-3750
- Turner, D., Sugimoto, N., and Friere, S. M. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 167<sup>5</sup>
   Krukowski, A. E., Chan, H. S., and Dill, K. A. (1995) J. Chem. Phys. 103,
- 10675 10688
- 14. Nozaki, Y., and Tanford, C. (1971) J. Biol. Chem. 246, 2211-2217
- 15. Wilce, M. C. J., Aguilar, M.-I., and Hearn, M. T. W. (1995) Anal. Chem. 67, 1210-1219
- 16. Sentell, K. B., and Henderson, A. N. (1991) Anal. Chim. Acta 246, 139-149
- 17. Schwarzenbach, R. P., Gschwend, P. M., and Imboden, D. M. (1993) Environmental Organic Chemistry, John Wiley & Sons, Inc., New York
- 18. van Gunsteren, W. F., King, P. M., and Mark, A. E. (1994) Q. Rev. Biophys. 27,  $435 - 481^{5}$
- 19. Suzuki, T., and Kudo, Y. (1990) J. Comput. Aided Mol. Des. 4, 155-198
- 20. Wyman, J., and Gill, S. J. (1990) Binding and Linkage: Functional Chemistry of Biological Macromolecules, University Science Books, Mill Valley
- 21. Ackers, G. K., Doyle, M. L., Myers, D., and Daugherty, M. A. (1992) Science 255, 54-63<sup>5</sup>
- 22. Ben-Naim, A. (1990) J. Chem. Phys. 93, 8196-8210
- 23. Sharp, K. A., Kumar, S., Rossky, P. J., Friedman, R. A., and Honig, B. (1996) J. Chem. Phys. 100, 14166-14177
- 24. Lumry, R., and Rajender, S. (1970) Biopolymers 9, 1125-1227<sup>5</sup>
- 25. Sturtevant, J. M. (1994) Curr. Opin. Struct. Biol. 4, 69-78<sup>5</sup>
- 26. Breslauer, K. J., Remeta, D. P., Chou, W. Y., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, J. G., and Markey, L. A. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8922-8926
- 27. Zimm, B. H., and Bragg, J. K. (1959) J. Chem. Phys. 31, 526-535
- 28. Poland, D. C., and Scheraga, H. A. (1970) Theory of the Helix-Coil Transition, Academic Press, New York
- 29. Flory, P. J. (1969) Statistical Mechanics of Chain Molecules, Wiley-Interscience, New York
- 30. Jacobson, H., and Stockmayer, W. H. (1950) J. Chem. Phys. 18, 1600
- 31. Flory, P. J. (1953) Principles of Polymer Chemistry, Cornell University Press, New York
- 32. Dill, K. A. (1985) Biochemistry 24, 1501-1509
- 33. Dill, K. A., and Stigter, D. (1995) Adv. Protein Chem. 46, 59-104<sup>5</sup>
- 34. Dill, K. A., and Shortle, D. (1991) Annu. Rev. Biochem. 60, 795-825<sup>5</sup>
- 35 Bromberg, S., and Dill, K. A. (1994) Protein Sci. 3, 997-1009
- 36. Chen, S.-J., and Dill, K. A. (1995) J. Chem. Phys. 103, 5802-5813
- 37. Miyazawa, S., and Jernigan, R. L. (1985) Macromolecules 18, 534-552
- 38. Sippl, M. J. (1990) J. Mol. Biol. 213, 859–883
- 39. Thomas, P. D., and Dill, K. A. (1996) J. Mol. Biol. 257, 457-469
- Page, M. I., and Jencks, W. P. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 40. 1678 - 1683
- 41. Chan, H. S., and Dill, K. A. (1994) J. Chem. Phys. 101, 7007-7026
- 42. Onsager, L. (1949) Ann. N. Y. Acad. Sci. 51, 627
- 43. deGennes, P. G. (1974) The Physics of Liquid Crystals, Clarendon Press, Oxford

<sup>5</sup> Review article.