EMPIRICAL FREE ENERGY CALCULATIONS OF PHAGE 434 REPRESSOR- AND *cro*-DNA COMPLEXES SUPPORT THE 'INDIRECT READOUT' HYPOTHESIS OF SPECIFICITY

LAWRENCE M. BROWN
Department of Molecular Biology, Princeton University, Princeton NJ 08544

ROBERT E. BRUCCOLERI and JIRI NOVOTNY
Department of Macromolecular Structure, Bristol-Myers Squibb, Princeton NJ 08543-4000

Empirical free energy calculations were done for two X-ray DNA-protein complexes. Complex formation energies were estimated for the phage 434 repressor and the 434 *cro* protein, both in complex with wild-type versions of their respective nucleic acid binding sites. The resulting quantities were subdivided according to the relative contributions of hydrophobicity, electrostatics and entropy (including both conformational changes and those pertinent to physical association of complex components). The non-linearized finite difference Poisson-Boltzmann equation was solved over a range of salt concentrations to determine magnitudes of coulombic and desolvation effects. Though approximate at best, results confirmed (1) the importance of centrally located charged protein residues that do not make direct contact with the DNA, (2) explicit protein backbone interactions with DNA, and (3) the varied parts of the two proteins used to harness the energy of binding. Although a large difference exists between calculated and experimentally measured free energy values, these can be reconciled by assuming that ≈ 1.5 e.u. (about 0.4-0.5 kcal at room temperature) of DNA vibrational and/or conformational entropy is lost for every basepair torsion in contact with the protein. Such a conclusion supports an "indirect readout" ("compound recognition") hypothesis of DNA binding specificity, wherein sequence-dependent differences in protein-DNA interaction are due to the topological peculiarities of the interactions rather than intrinsic chemical differences alone.

Introduction

Empirical free energy calculations have been used successfully to delineate contributions to specificity in protein-protein complexes to the atomic level (reviewed in, e.g., Vajda et al. 1997, Ajay & Murcko 1995). For example, the idea of an "energetic" or "functional" epitope, first postulated from such empirical estimates (Novotny et al. 1989) was subsequently confirmed by site-directed and alanine-scanning mutagenesis of a number of protein-protein systems (e.g., Jin et al. 1992). In this communication, we report on our first attempt to estimate the Gibbs free energy of association of the bacteriophage 434 and *cro* proteins with their cognate DNA oligomers. The X-ray crystallographic structures of these two complexes (Mondragon & Harrison 1991, Shimon & Harrison 1993) as well as the approximate free

energy of association (Koudelka et al. 1987, Takeda et al. 1992) have been known for some time. These two structures were selected because great conformational changes were undergone by neither the protein nor the DNA (Harrison et al 1988), hence providing an example of an approximately rigid, "lock-and-key" type of macromolecular association (Harrison et al. 1988).

From the technical point of view, electrostatic binding energy calculations on highly charged (e.g., DNA) molecules present special difficulties (cf., e.g., Sharp & Honig 1990). This report gives a brief description of the Gibbs free energy function used, the conditions necessary for the solution of the non-linear Poisson-Boltzmann equation, and the results of said calculations, with emphasis placed on the aforementioned "indirect readout" hypothesis.

Methods

The AMBER94 parameters and partial atomic charges (Cornell et al. 1995; see also, e.g., Cheatham & Kollman 1996, 1997) were used throughout our calculations.

The formal concepts of free energy attribution arising from noncovalent complex formation remain the same as before (Novotny et al. 1989,1997). Gibbs free energy (or ΔG) is considered to be comprised of hydrophobic effect, electrostatic interactions, and entropic changes (either conformational changes in the solute, or changes in association between solute components or solvent with solute).

In the most current of our communications (Novotny et al. 1997), our free energy function's form proposes that the *hydrophobic ''force''* (solvent entropy differences associated with a change of solvent/solute interface), ΔG_{HB}, is directly proportional to the size of the said interface, as defined by its molecular surface, and hence:

$$\Delta G_{HB} = 0.07 \times (MolecularSurfaceDifference)\ [kcal];$$

(the hydrophobic scaling constant, 70 cal/mol/Å^2, is that suggested by Tunon et al. in 1992). The program CONGEN (CONformational GENerator) employs the GEPOL (Pascual-Ahuir & Silla 1990) algorithm to construct the molecular surface. These were assembled into CONGEN ANALYSIS tables in conjunction with corresponding residue and peptide chain sums.

Conformational entropy changes or $T\Delta_{CF}$ were estimated by the simplified formula

$$T\Delta S_{CF} = RT\ln(\frac{N_{free}}{N_{bound}})$$

where R is the gas constant, 1.987 cal/mol; T is temperature in degrees Kelvin; N_{free} and N_{bound} are the number of torsions available in the free and the bound state, respectively. The numbers of available conformations, N_{free} and N_{bound}, were enumerated by CONGEN conformational searches over the preferred side chain rotamers.

The *electrostatic effects* were obtained from the equation $\Delta G = 1/2\int\rho\phi(r)$, ($\phi(r)$, electrostatic potential as a function of position). These calculated effects include the formation and breaking of hydrogen bonds, the solvation and desolvation of charges, and pure coulombic effects. In our analysis, these were obtained from the non-linear Poisson-Boltzmann equation or,

$$-\nabla\cdot\varepsilon(r)\nabla\phi(r)+\kappa^2[\phi(r)] = 4\pi\rho(r)$$

(wherein $\varepsilon(r)$ is the dielectric constant as a function of position, r; κ is the Debye-Huckel constant; ρ is electric charge density). The finite difference approximation of the Poisson-Boltzmann equation, FDPB, assigns molecules to a three-dimensional rectangular grid, with atomic point charges, and dielectric constant values ($\varepsilon = 78$ for the solution, $\varepsilon = 4$ inside the solute), distributed on the grid points nearest to their actual locations. The field equations are then solved for the values of the electrostatic potential for each grid point, by numerical iteration over the very many finite difference equations describing field differences between the 6 pairs of orthogonal, neighboring grid points, i and j. The equation employed is:

$$\kappa_i^2\phi_i+\sum_{j=1}^{6}\varepsilon_{ij}(\phi_i - \phi_j) = \rho_i h^2$$

(h, point-to-point distance, i.e., grid spacing). CONGEN's use of the FDPB algorithm features adjustable rectangular grids suitable for macromolecular calculations, a scheme for the uniform assignment of charge such that unfavorable grid energies are avoided, and smoothing algorithms that alleviate problems associated with the highly irregular nature of macromolecular dielectric boundaries and/or the discretization of space arising from the gridding scheme. It has been known that the infinite self-energy of a point charge embedded in the Maxwell electrostatic field becomes, in the FDPB algorithm, a grid size-dependent component of the

potential (the so called "grid energy"). This energy tends to grow to very large values as the grid itself becomes smaller (i.e., commensurate with, or smaller than, atomic dimensions, or approximately 1 Å). Therefore, the paradox lies in the fact that smaller grids, although they better approximate the details of molecular shapes, have proportionally larger energies, and hence comparably larger errors when positions of molecules on the grid are varied. This makes subtractive calculation:

$$\Delta G_{EL,binding} = \Delta G_{EL,complex} - \Delta G_{EL,protein} - \Delta G_{EL,DNA}.$$

with sufficient numerical precision difficult. As communicated before (Bruccoleri 1993, Bruccoleri et al. 1997) the uniform charge assignment scheme, harmonic averaging of dielectric points and charge anti-aliasing have greatly reduced problems with grid energies, as well as those associated with the steepness, shape and discretization of dielectric boundaries.

The calculations with the FDPB were done with the CONGEN program using 0.8 Å grid, with uniform charging (Bruccoleri 1993), anti-aliasing and 15-point harmonic smoothing. A 8 Å grid margin was maintained around the molecules. Dielectric constant of the solvent was set to 78, that of the solute to 4. Calculations were carried out at room temperature (300° K) at the solvent ionic strength = 0.15. Molecular surfaces were calculated using the GEPOL algorithm as implemented in CONGEN; the sphere overlap factor was set to 0.7. A single electrostatic energy determination consisted of 10 independent subtractive calculations, each with a different relative position of the molecules on the spatial grid. Positional error on the average, was approximately 1% of the calculated value (Table 5). A single subtractive calculation consisted from 3 electrostatic potential evaluations, all on the same grid: (i) the complete DNA-protein complex, (ii) the protein alone, and (iii) the DNA alone.

Due to the high charge density of the phosphodiester backbone of the nucleic acid duplex, the finite difference Poisson-Boltzmann equation had to be solved in its non-linear form. The nonlinear regime for computation often converges slowly or not at all. After a series of trial calculations aimed at identifying the optimal parameters for good solutions we employed a very large maximum iteration limit (40,000 iterations) with the convergence (tolerance) limit 0.00001. The under-relaxation parameter was set to 0.3. With our hardware, an SGI 24-processor ONYX workstation with 2 GB of RAM, one set of binding energy calculations typically completed in 12 to 24 hours.

Results

In our empirical scheme, individual binding energy terms are calculated as atomic properties that are later summed into tables and presented as the individual contributions of residues and molecules.

Table 1. Calculated Gibbs free energies

CALCULATED COMPONENT	PROTEIN	
	434 Repressor	**434 CRO**
Contact molecular surface [Å^2]	2104	2273
ΔG_{HB} [kcal]	-147.3	-159.1
ΔG_{EL} [kcal]	12.1	-3.8
$T\Delta S_{CF}$ [kcal]	2.8	3.9
$T\Delta S_{AS}$ (estimate, see Finkelstein & Janin 1989)	15	15
ΔG_{TOTAL} [kcal]	$\approx$ -60†	$\approx$ -69†
$\Delta G_{EXPERIMENTAL}$ [kcal]	**-10.3**	**-9.5**

†Calculated as $\Delta G = 0.6\times\Delta G_{HB} + 0.4\times\Delta G_{EL} + 1.3\times(T\Delta S_{CF} + T\Delta S_{AS})$, (see Novotny et al. 1997).

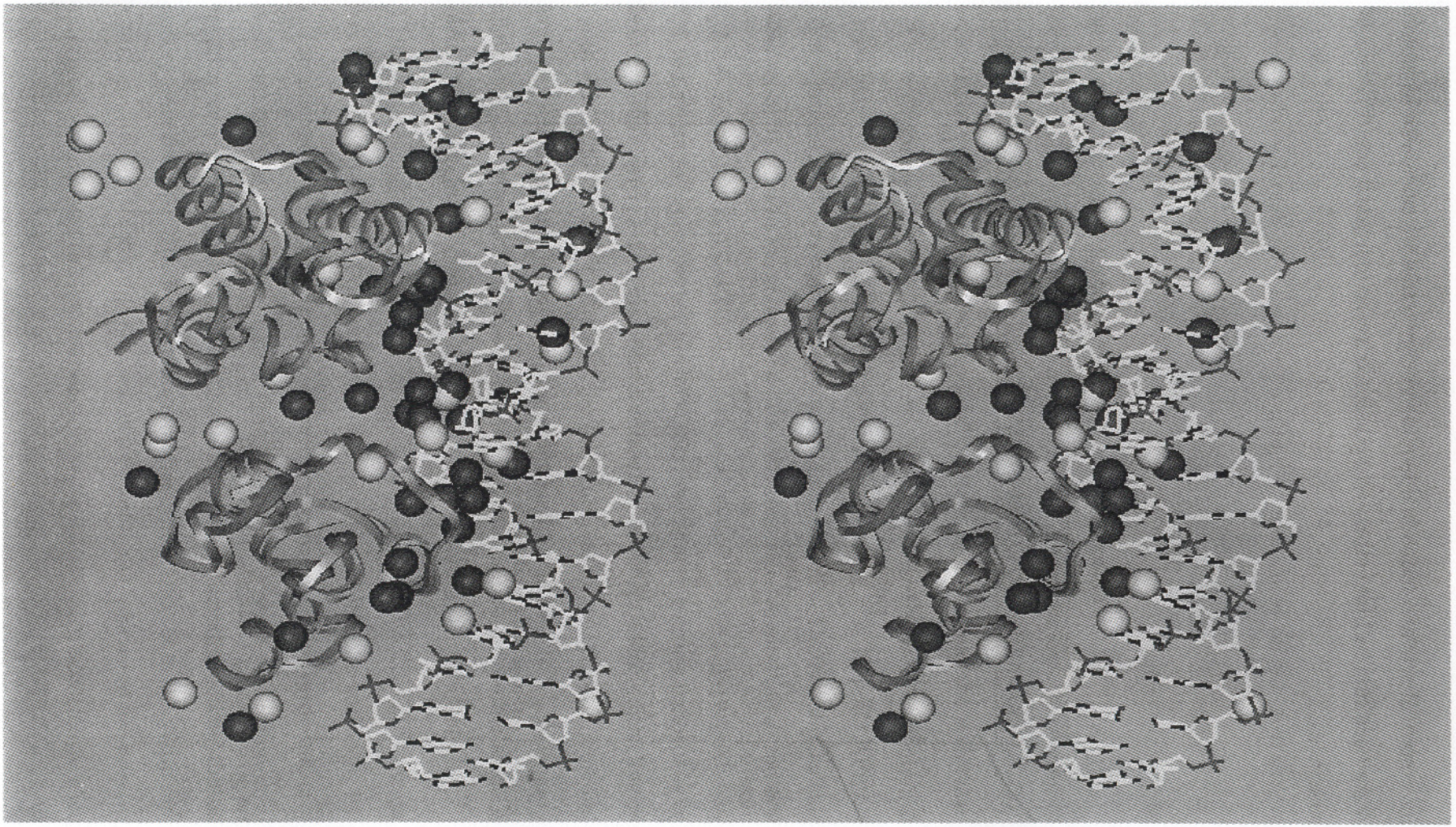

Figure 1. Least–squares superposition of the bacteriophage 434 repressor–DNA and cro–DNA complexes with crystallographically resolved water molecules shown as van der Waals spheres

Table 1 summarizes the hydrophobic, electrostatic and entropic contributions of the molecules subjected to the computational analysis. Large protein-DNA contact areas, and their correspondingly large hydrophobic effects, dominate the energetics. We must note that the electrostatic contribution is relatively small in comparison. The calculations, as noted in Table 2, were carried out for various ionic strengths including 0.018 M salt, i.e. the salt concentration used in the experimental binding constant determinations.

Both complexes contained crystallographically resolved water molecules that were, as a rule, found at different locations in the two different complexes (Figure 1). In particular, number of the water molecules localized in the major groove of the repressor DNA was not clearly visible in the *cro* structure. In view of these difficulties, and in the absence of controverting data, we believe that the continuous model of hydration, as implicitly adopted in our formulas, represents a reasonable and convenient initial approximation to the energetics of DNA-protein binding.

An examination of the energetically most important repressor and *cro* residues uncovered similarities as well as differences in the way these two proteins engaged their cognate DNA sites. Approximate energetics of repressor binding, dissected into individual side chain contributions, pointed out important roles of the several charged arginine side chains engaged in both electrostatic and large contact area (hydrophobic) interactions with the DNA molecules.

Discussion.

The numerical precision of our FDPB solutions (results within a fraction of a kilocalorie, Novotny et al. 1997) has been achieved in this calculation as well (Table 2). One remaining source of uncertainty lies with the presence of water at the protein/DNA interface, a phenomenon that has not been treated explicitly. On this point the crystallographic data seem to be inconsistent (there is no established or uniform pattern of DNA solvation in the two crystal structures) and the theoretical framework of binding energy calculations with explicit water molecules is only now being formulated (cf., e.g., Novotny & Bajorath 1996). It is our belief that the continuum model used in these computations is the best possible approximation presently available.

Table 2. Calculated electrostatic binding energies for bacteriphage 434 repressor-DNA and CRO-DNA at different salt concentrations

IONIC STRENGTH	ΔG_{EL} [kcal]	
	repressor-DNA	CRO-DNA
0.000	−31.9±0.4	−64.8±0.8
0.018	12.1±0.4	−3.8±0.7
0.100	19.1±0.4	5.7±0.6
0.500	24.1±0.4	12.2±0.6
1.000	25.8±0.4	14.4±0.6

The most conspicuous feature of Table 1 is the great discrepancy between the calculated and experimentally determined values of Gibbs free energies. Although past work achieved accuracy of the $\Delta\Delta G$ calculation within about 1.3 kcal (Novotny et al. 1997), the current error appears to be much larger (some 50 kcal or, ≈ 500%). One possible explanation for the missing energy can be found in the "indirect readout" hypothesis (Draper 1993, Spolar & Record 1994). Recent experimental evidence (Lesser et al. 1993) and thermodynamic measurements (Takeda et al. 1992) seem to support the postulate that, upon protein binding, concerted sequence-specific changes in DNA structure occur that optimize binding specificity. It than follows that some of the free energy of binding would have to be expended on conformational changes in the DNA molecule. The conformational-entropic content of DNA structure in solution is poorly understood but is assumed to be rather large. It is known, however, that there are 10 torsional degrees of freedom for every single basepair in the forms of 2' and 3' endo- and exo-deoxyribose puckers, two torsions each on the 5' and 3' O-side of the phosphate group, and the glycosidic bond which covalently links the ribose sugar and the pyrimidine or purine bases. Assuming, for the sake of simplicity, that each torsion oscillates in solution between no more than two distinct yet equienergetic states, a complete immobilization of ten 2-state rotators/vibrators would carry an approximate conformational entropy

cost of $10 \times [RT\ln(2/1)] \approx 4$ kcal. As 13 basepairs are found in intimate contact with the protein in both the repressor and *cro* complexes, as much as 52 kcal of free energy of binding may become lost upon binding.

Perhaps a more realistic picture of DNA flexibility would be that of more global conformational shifts spreading beyond the basepairs in immediate contact with the protein, and only partially immobilized, but with a comparable conformational entropy penalty. Nevertheless, the conjectural 52 kcal of energy cost is commensurate with our missing free energy, i.e., 50 kcal for the repressor and 59 kcal for the *cro* protein (cf. Table 1).

In our opinion, the only alternative explanation for the large excess of calculated free energies of binding could be found in possible destabilizing effects of ≈ 30 water molecules passively trapped at the protein-DNA interface. Simple estimates of water energetics at macromolecular interface (Dunitz 1994; Novotny & Bajorath 1996) do point to mostly destabilizing or, approximately at best, neutral contributions. In this case, the unfavorable translational/rotational entropy change of water immobilization (≈ 2 kcal) is partly compensated for by intermolecular, water/protein and/or water/DNA, hydrogen bonds. Yet, in our case, the probability that many destabilizing water molecules are trapped passively at an intermolecular interface so heavily populated with formally charged groups seems small. In the situation of such an electron-rich environment, one must expect water to be engaged in hydrogen-bonding networks that are, in general, energetically favorable (Novotny et al. 1990, Rejto & Verkivker 1997). This type of hydration would imply water molecules to be tightly bound to the oligonucleotide prior to complex formation (Shakked et al. 1994) and therefore entering into the bimolecular binding reaction: protein - hydrated DNA, with no translational/rotational entropy penalties.

In summary, the current study represents our first approximate account of binding energetics of protein-DNA complexes. In our view, the results point towards the possibility of large DNA conformational entropy shifts on protein ligation, however, more computational and crystallographic work will be needed for final resolution of all the issues.

References

Ajay & Murcko, M. A. (1995) *J. Med. Chem.* **38,** 4953-4967.

Bruccoleri R. E. (1993) *J. Comp. Chem.* **14,** 1417-1422.

Bruccoleri R. E., Novotny J., Davis M. & Sharp K. A. (1997) *J. Comp. Chem.* **18,** 268-276.

Cheatham T. E. & Kollman P. A. (1996) *J. Mol. Biol.* **259,** 434-444.

Cheatham T. E. & Kollman P. A. (1997) *J. Am. Chem. Soc.* **119,** 4805-4825.

Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Furguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. (1995) *J. Am. Chem. Soc.* **117,** 5179-5197.

Draper D. E. (1993) *Proc. natl. Acad. Sci. U.S.A.* **90,** 7429-7430.

Dunitz J. D. (1994) *Science* **264,** 670.

Finkelstein A. V. & Janin J. (1989) *Protein Eng.* **3,** 1-3.

Harrison S. C., Anderson J. E., Koudelka G. B., Mondragon A., Subbiah S., Wharton R. P., Wolberger C. & Ptashne M. (1988) *Biophys. Chem.* **29,** 31-37.

Jin L., Fendly B. M. & Wells J. A. (1992) *J. Mol. Biol.* **226,** 851-865.

Koudelka G. B., Harrison S. C., Ptashne M. (1987) *Nature* **326,** 886-891.

Lesser D. R., Kurpiewski M. R. Waters T., Connolly B. A. & Jen-Jancobson L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 7548-7552.

Mondragon A. & Harrison S. C. (1991) *J. Mol. Biol.* **205**, 179-190.

Novotny J., Bruccoleri R. E. & Saul F. A. (1989) *Biochemistry* **28,** 4753-4749

Novotny J. & Bajorath J. (1996) *Adv. Protein Chem.* **49,** 150-260

Novotny J., Bruccoleri R. E., Davis M. & Sharp K. A. (1997) *J. Mol. Biol.* **268,** 401-411

Pascual-Ahuir J. L. & Silla E. (1990) *J. Comp. Chem.* **11,** 1047-1060.

Sharp K. A. & Honig B. (1990) *J. Phys. Chem.* **94,** 7684-7692.

Shimon L. W. & Harrison S. C. (1993) *J. Mol. Biol.* **232** 826-841.

Spolar R. S. & Record M. T. (1994) *Science* **263,** 777-784.

Takeda Y., Ross P. D., Mudd C. P. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89,** 8180-8184.

Tunon I., Silla E. & Pascual-Ahuir J. L. (1992) *Protein Eng.* **5,** 715-716.

Vajda S., Novotny J. & Sippl M. (1997) *Cur. Opinion Struct. Biol.* **7,** 222-228.