From a practical standpoint, current clinical trials are evaluating the ability of ARIs either to prevent the progression or reverse the presence of clinically observed pathology. A major anticipated problem with this approach, however, is that the timing of intervention may not be optimal. Checking the progress of diabetic pathology may not be possible once the pathological process has progressed to a certain point. Moreover, the dosage of inhibitor at that point may not be adequate. Thus, despite the fact that aldose reductase may be involved in the initiation of the pathology, the clinical trials with ARIs may not demonstrate this point. Intervention studies conducted on cataract formation in galactosemic rats suggests such a possibility.[36] While cataract formation can be essentially prevented through administration of ARI concomitant with the galactose diet, the "reversal" of the cataractogenic process can only be accomplished through either the removal of the high galactose diet or the administration of an adequate dose of ARI prior to the sixth day of galactose feeding. After the sixth day a point of no return was reached in the reversal of the cataractous process with the vacuolar stage continuing on to the dense nuclear opacity.

The galactose cataract study indicates the required administration of ARI at the initial onset of the pathogenic process, and to date, most animal studies with ARIs have stressed prevention of diabetic pathology rather than intervention. These animal studies suggest that ARIs should eventually be clinically administered at the onset of diagnosed diabetes on a prophylactic basis. For such long-term administration ARIs with extremely low toxicity and no side effects will be needed.

**Acknowledgment.** We thank Drs. Dennis Feller and Herman Ziffer of the National Institutes of Health for their helpful comments and discussions.

**Registry No.** Aldose reductase, 9028-31-3.

(94) Judzewitsch, R.; et al. *Diabetes* **1981**, *30* (suppl 1), 30A.

# *Articles*

## A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules

P. J. Goodford

*The Laboratory of Molecular Biophysics, The Rex Richards Building, University of Oxford, Oxford OX1 3QU, England.*
*Received August 3, 1984*

The interaction of a probe group with a protein of known structure is computed at sample positions throughout and around the macromolecule, giving an array of energy values. The probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl. Contour surfaces at appropriate energy levels are calculated for each probe and displayed by computer graphics together with the protein structure. Contours at negative energy levels delineate regions of attraction between probe and protein and are found at known ligand binding clefts in particular. The contours also enable other regions of attraction to be identified and facilitate the interpretation of protein–ligand energetics. They may, therefore, be of value for drug design.

Several attempts have been made to use the atomic structures of biological macromolecules in order to discover novel therapeutic agents.[1] The novel molecules are designed so that they should fit onto an appropriate protein whose structure has been previously determined by X-ray crystallography. However, this fitting process is not just a matter of simple geometry because the chemical properties of the individual atoms are critically important, and the goodness of fit may be assessed by predicting the energetics of the drug–protein system instead of its shape.[2] A method is now described in which energy contour surfaces are displayed in three dimensions on a computer graphics system together with the macromolecular structure, so that energy and shape can be considered simultaneously when designing drugs.

The idea of small hard atoms goes back to Democritus and Lucretius, but it was van der Waals who used the Joule–Thompson effect to show that there are attractive forces between atoms. The combination of attractive with repulsive forces has led to a number of simple functions[3,4] that can display energy minima corresponding to preferred nonbonded arrangements of atoms and molecules, and the (12,6) Lennard–Jones potential (Figure 1) is such a function:

$$E_{lj} = A/d^{12} - B/d^{6} \quad (1)$$

In this equation $d$ is the distance between a pair of nonbonded atoms whose Lennard–Jones energy $E_{lj}$ is described by the parameters $A$ and $B$. When $d$ is small, the $A/d^{12}$ term generates a dominating repulsion corresponding to a large positive value of $E_{lj}$. This effectively defines a minimum separation that can be apportioned between the atoms, giving each of them a nominal radius, and the threshold step function established in this way can be extended to larger systems, thus determining a molecular surface that may be measured[5] or displayed by computer graphics.[6] Alternatively, a somewhat bigger value of $d$ may be used, corresponding to the minimum value of $E_{lj}$ at the

(1) Goodford, P. J. *J. Med. Chem.* **1984**, *27*, 557.
(2) Richards, W. G. "Quantum Pharmacology", 2nd ed.; Butterworths: London, 1983.
(3) Buckingham, R. A. *Proc. R. Soc. London, A* **1938**, *168*, 264.
(4) Jones, J. E. *Proc. R. Soc. London, A* **1924**, *106*, 463.
(5) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(6) Blaney, J. M.; Jorgensen, E. C.; Connolly, M. L. Ferrin, T. C.; Langridge, R.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J. Med. Chem.* **1982**, *25*, 785.

0022-2623/85/1828-0849$01.50/0 

**Table I.** Parameters Used To Evaluate the Nonbonded Interactions of Probe Groups (See Text)[a,b]

| probe | | Lennard–Jones | | | | hydrogen bond | | | |
|---|---|---|---|---|---|---|---|---|---|
| description | symbol | $\alpha_i$ | $N_{eff}$ | rad | charge | $C$ | $D$ | H→ | H← |
| amino | $NH_3^+$ | 2.13 | 9 | 1.75 | 0.66 | 4295 | 629 | 3 | 0 |
| carbonyl oxygen | O | 0.84 | 6 | 1.60 | −0.37 | 3855 | 738 | 0 | 2 |
| carboxy oxygen | $O^-$ | 2.14 | 6 | 1.60 | −0.57 | 3855 | 738 | 0 | 2 |
| hydroxyl | OH | 1.20 | 7 | 1.65 | −0.12 | 3855 | 738 | 1 | 1 |
| methyl | $CH_3$ | 2.17 | 8 | 1.95 | 0.00 | | | 0 | 0 |
| water | $H_2O$ | 1.20 | 7 | 1.70 | 0.00 | 3855 | 738 | 2 | 2 |

[a] H→ and H← are the maximum number of hydrogen bonds that each probe can donate and accept. [b] Units: polarizability ($\alpha_i$), Å$^3$; van der Waals radius (rad), Å; charge, the electron charge; $C$, kcal mol$^{-1}$ Å$^6$; $D$, kcal mole$^{-1}$ Å$^4$.

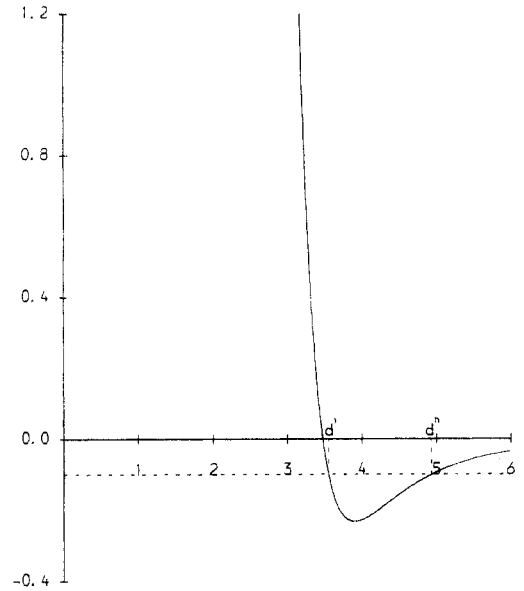



**Figure 1.** A Lennard–Jones curve for the interaction between two nonbonded atoms. Abscissa: interaction energy $E_{lj}$. Ordinate: interatomic distance $d$. Slightly negative values of $E_{lj}$ yield two alternative distances $d'$ and $d''$, which define a zone of interatomic attraction, and some characteristics of such zones have now been examined by computer graphics. See text.

bottom of the trough in the Lennard–Jones curve, and this generates a rather more extended molecular surface for measurement or display.

An attempt has now been made to study the surfaces characterized by a slightly negative energy value that defines two intersections with the Lennard–Jones curve (Figure 1) at distances $d'$ and $d''$. As atoms approach nearer together than $d'$, the repulsion term $A/d^{12}$ soon dominates, while the interaction tends to fade away completely as the atoms move further apart than $d''$. There is a zone of mutual attraction between these limits, and some characteristics of the three-dimensional contour surfaces at $d'$ and $d''$ that define the extent of the zone in macromolecular systems are now explored. In order to do this the interaction of a probe group with the chosen protein is computed at sample positions throughout and around the macromolecule, giving an array of energy values through which the contours may be drawn. Probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl.

Electrostatic and hydrogen bond effects are also considered, and the present studies show that some known ligands do bind as expected in the calculated zone of attraction. It is suggested that the contour surfaces at $d'$ and $d''$ may be of value in designing novel ligands or drugs.

## Methods

The overall aim is to compute an empirical energy function and thereby produce an array of energy values. Three-dimensional contour surfaces at selected energy levels may then be generated and studied together with the protein structure. The examples given are displayed on an Evans and Sutherland Picture System 2 using program FRODO,[7,8] but these display procedures are well established and it is only necessary to describe the earlier stages of the method. Moreover, the results can be displayed effectively on other graphics systems.

**Preparation of Data.** Atomic coordinates for the chosen macromolecule are taken from a source such as the Protein Data Bank[9] and should be carefully checked against the notes and literature references. An appropriate choice is made where alternative atomic positions have been provided, and insertions may be supplied for missing groups. Ligands such as water molecules, chloride ions, or cofactors are not normally treated as part of the macromolecule, unless they are structurally significant like the heme group of hemoglobin.

A table of parameters (table GRUB) is needed in order to evaluate the Lennard–Jones and other empirical energy functions, and program GRIN appends these parameters to the atomic coordinates of the protein. Each "heavy" atom in each of the 20 common amino acids is individually considered, and additional molecules such as hydroxyproline or heme may be added to the table and processed in the same way. It may then be appropriate to modify some of the parameter values in particular cases. Thus the electric charges on amino and oxy terminal atoms and disulfide bridge atoms may be changed as suggested by McCammon et al.,[10] and account may be taken of the p$K_a$ of histidine residues if this is known. Hydrogen-bonding hydrogen atoms in positions of standard geometry are appended to the protein by program GRIN if they are not already present, and a check is made on Cys sulfur atoms to be sure they have received proper treatment. For proteins such as hemoglobin, the complete physiological oligomer must then be generated by symmetry operations, and a check is finally made for atoms that may be wrongly listed twice in the same position on symmetry axes.

The parameters in GRUB are based on the "extended" atom concept[10,11] used for program CHARMM. Thus, the amino group is treated as a single entity with a van der Waals radius of 1.75 Å, which is somewhat larger than the normal value for a nitrogen atom. This single "extended" amino group replaces four real atoms for computational purposes, so that the size and duration of all the computations is significantly reduced. Other advantages and disadvantages of the "extended" atom representation have been considered by Brooks et al.,[11] and the parameter

(7) Jones, T. A. *J. Appl. Crystallogr.* **1978**, *11*, 268.
(8) Jones, T. A. "Computational Crystallography"; Sayre, D., Ed.; Clarendon Press: Oxford, 1982; p 303.
(9) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Bryce, M. D.; Rodgers, J. R.; Kennard, O.; Shikanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(10) McCammon, J. A.; Wolynes, P. G.; Karplus, M. *Biochemistry* **1979**, *18*, 927.

values for "extended" probe groups are presented in Table I.

**The Grid.** The calculations are performed by Fortran program GRID, which can be made available to interested parties together with program GRIN and table GRUB. A regular array of "GRID points" is established throughout and around the protein, and the potential energy $E_{xyz}$ of the probe is calculated when it is located at the first point on the first $XY$ plane of the GRID. Successive probe positions are sampled in the same way until an energy value $E_{xyz}$ has been assigned to every GRID point.

The dimensions of the array are determined so that all points on the first $XY$ plane are outside the protein, and the computed energy values are therefore small when the probe is in this plane. However, subsequent planes start to intersect the macromolecule, and large positive energies due to the $A/d^{12}$ Lennard–Jones repulsion term may then be calculated for any GRID point that happens to be near a protein atom. Other points lie in the interatomic spaces, and modest negative energies would then correspond to favorable interactions between the probe and the protein. These would be partly due to the $-B/d^6$ term of the Lennard–Jones function, partly to electrostatic effects, and partly to hydrogen-bond interactions. Further arrays may also be calculated for other probes of interest, and each may be separately contoured for display so that the interactions of different ligand types with the same protein can be compared.

**The Empirical Energy Functions.** The nonbonded interaction energy $E_{xyz}$ of the probe at each $xyz$ position on the GRID is calculated as the sum of many different components:

$$E_{xyz} = \sum E_{lj} + \sum E_{el} + \sum E_{hb} \qquad (2)$$

Each individual term in the summations relates to one pairwise interaction between the probe at position $xyz$ and a single "extended" atom of the protein. The summations extend over all "extended" protein atoms, although some small terms may be set to zero as indicated below in order to save computing time.

**The Lennard–Jones Function.** The form of the Lennard–Jones terms $E_{lj}$ has already been defined (eq 1), but $E_{lj}$ is set to zero if the probe and the pairwise protein atom are more than a certain distance apart. This is currently 8 Å at which $E_{lj}$ is typically −0.01 kcal/mol and is, moreover, diminishing according to the sixth power of the interatomic distance. The cut-off at 8 Å yields a worthwhile saving of computer time, although the neglect of small terms may introduce an error of the order of 0.1 kcal/mol in $\sum E_{lj}$.

The values of $A$ and $B$ in eq 1 are calculated from the effective number of electrons $N_{eff}$, the polarizability $\alpha_1$, and the van der Waals radius $r$ of the interacting atoms as described by Hopfinger.[12]

**The Electrostatic Function.** The electrostatic interaction $E_{el}$ does not diminish rapidly with distance, and the individual terms are therefore evaluated pairwise between the probe group and every "extended" protein atom, irrespective of their separation. However, the magnitude of $E_{el}$ is critically sensitive to the spatial dielectric behavior of the environment, which has been considered by Hopfinger,[12] who proposed a distance-dependent dielectric. In the present work it is assumed that a planar interface separates a homogeneous protein phase of dielectric $\zeta$ from a homogeneous solution of dielectric $\epsilon$. The nominal depth ($s_q$) of each protein atom in the protein phase is assessed by counting the number ($n$) of neighboring protein atoms whose nuclei lie within a distance of 4 Å, and the depth ($s_p$) of the probe at each $xyz$ position is similarly assessed. In order to calibrate this procedure, a test sphere of radius 4 Å was placed at various positions in or around a protein during a series of preliminary studies, and the number of protein atoms in the sphere was counted. When it contained 12 or more atoms, the sphere was apparently immersed in the protein phase, but if it held less than seven atoms, the center of the sphere appeared to be in the solution. Table II shows the depth ($s$) of the sphere's center into the protein phase for other numbers of atoms on the basis of these studies, and $E_{el}$ may then be evaluated directly by applying the method of images[13-15]

**Table II.** Assessment of the Nominal Depth ($s$) of a Probe Group or Atom inside the Protein, Which Is Regarded as a Homogeneous Phase (See Text)[a]

| $n$ | $s$, Å | $n$ | $s$, Å |
|---|---|---|---|
| ≤6 | 0.0 | 10 | 1.9 |
| 7 | 0.4 | 11 | 2.6 |
| 8 | 0.9 | ≥12 | 4.0 |
| 9 | 1.4 | | |

[a] $n$ is the number of protein atoms within 4 Å of the probe position.

$$E_{el} = \frac{pq}{K\zeta}\left[\frac{1}{d} + \frac{(\zeta - \epsilon)/(\zeta + \epsilon)}{\sqrt{d^2 + 4s_p s_q}}\right] \qquad (3)$$

In this equation $p$ and $q$ are the electrostatic charges on the probe group and the pairwise protein atom that are separated by a distance $d$, and $K$ is a combination of geometrical factors and natural constants.[11] The term $4s_p s_q$ is set to zero if the probe has less than seven protein neighbors within a distance of 4 Å, since this gives the appropriate functional form on the present assumptions according to the method of images when the probe is outside the protein phase. Values of 80 and 4 were used for $\epsilon$ and $\zeta$.

Equation 3 represents a compromise between procedures based upon the algorithm of Warwicker and Watson,[15] which can require substantial computing times, and the classical functions whose inadequacies have been discussed by Hopfinger.[12] Equation 3 leads to an effective dielectric value of $\zeta$ if the pairwise groups are so deep in the protein and so close together that the effect of the solution may be neglected. At increasing separations there is a distance dependency analogous to that suggested by Hopfinger,[12] but if one or both groups approach the surface the electrostatic interaction becomes progressively diminished, and it is almost switched off as the probe enters the solution when the effective dielectric rises to $(\zeta + \epsilon)/2$. Some of the constants in eq 3 can be collected in order to speed up the computation, and it is only necessary to determine $s_p$ and $s_q$ once in order to evaluate eq 3 for every interaction. The square-root term is rather time consuming, but it does not have to be evaluated when the probe is outside the protein.

**The Hydrogen Bond Function.** A direction-dependent 6–4 function is used for hydrogen bonds (eq 4), giving

(11) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swiminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.
(12) Hopfinger, A. J. "Conformational Properties of Macromolecules"; Academic Press: New York, 1973; Chapter 2.
(13) Landau, L. D.; Lifshitz, E. M. "Course of Theoretical Physics"; Englis ed.; Pergamon Press: Oxford, 1960; Vol. 8.
(14) Rogers, N.; Sternberg, M. *J. Mol. Biol.* **1984**, *174*, 527.
(15) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.

a broad energy minimum as suggested by Brooks et al.[11]

$$E_{hb} = [C/d^6 - D/d^4] \cos^m \theta \quad (4)$$

When two identical atoms are interacting, the tabulated values for $C$ and $D$ determine their interatomic separation $d_{min}$ at the bottom of the curve. If the atoms are of different types, the geometric mean of their individual $D$ values and the arithmetic mean of their $d_{min}$ separations are used. The appropriate $C$ value is then calculated by setting the differentiated form of eq 4 equal to zero, which leads to

$$C = 0.666 D d_{min}^2 \quad (5)$$

If the protein donates a hydrogen bond, then the bond direction is determined by the hydrogen position as computed from the heavy atom structure of the protein. $\theta$ is the angle $DHP$ where $D$ is the protein donor atom, $H$ is the hydrogen, and $P$ is the probe accepting the hydrogen bond. The term $m$ is normally 4, but the whole $E_{hb}$ term is set to zero when $\theta \leq 90°$. However, problems can arise when tyrosine donates a hydrogen bond, because it is not always easy to decide on an unequivocal position for the hydrogen atom. In the present work a dummy hydrogen position is assigned on the tyrosine side-chain axis at 1 Å from the phenolic oxygen, and a less well-defined orientation is specified for this hydrogen bond by setting $m$ = 2. The procedure tends to overestimate the strength of hydrogen bonds donated by tyrosine, but the bias is too small to show on any of the present figures.

If the probe group donates the bond, it is assumed that the probe can orient itself in order to form the most effective hydrogen-bond interaction with the acceptor atom of the protein, and the cos $\theta$ term is set to unity.

If $E_{lj}$ indicates a Lennard–Jones repulsion but $E_{hb}$ shows a favorable hydrogen-bond interaction between two atoms, then $E_{lj}$ is set to zero. The maximum number of hydrogen bonds that each protein atom can donate and the maximum number that it can accept are specified and are also specified for the probe. Out of the many possible hydrogen bonds that a probe might be able to make at any particular $xyz$ position, the energetically most favorable ones are then selected subject to these constraints.

There is no general agreement at present on the empirical functions that best describe $E_{hb}$ or, indeed, $E_{lj}$ and $E_{el}$. Some authors[11] favor the use of trigonometrical functions for hydrogen bonds, but others[16-18] suggest that directional preferences can be adequately simulated without explicit angular factors. Further work is needed, and no particular claim is made for the present functions and parameters in preference to those of Snir et al.[19] or Weiner et al.[20]

## Results

**Water Contours.** The full parameterization of water presents particular problems, and a very simple model has therefore been used. An "extended" water molecule ($H_2O$; Table I) is treated as an electrically neutral group that has no dipole but can donate up to two hydrogen bonds and can also accept up to two. It also seemed desirable to use a sufficiently close GRID spacing between GRID points to ensure that no appreciable cavity in the matrix of the protein would be passed over, and with this in mind the first computations were performed with water as a probe group at a GRID spacing of 0.5 Å. The three-dimensional contours in Figure 2 are therefore displayed as a network of interlocking vectors whose lengths correspond to this distance.

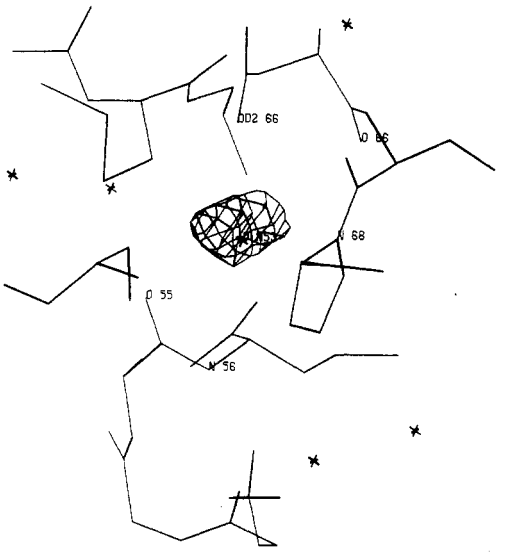



**Figure 2.** The cluster of short vectors were calculated from the observed structure of the phospholipase-A2 macromolecule.[21] These vectors define a surface that surrounds a globular region in a cavity in the protein. A water molecule in this region should have an interaction energy of −9 kcal/mol with the enzyme, according to the present calculations. This corresponds to a significant attraction, and a water molecule (H53) was observed by X-ray crystallography as shown. It is almost exactly where predicted by the contours.

Part of the bovine phospholipase-A2 macromolecule[21] is depicted in Figure 2 with contours plotted at a height of −9 kcal/mol. This contour level corresponds to a significant attraction, and the binding of a water molecule should be well favored in the small globular region defined by the contours. The size of this zone is largely determined by Lennard–Jones interactions while its suitability for water is due to the presence of nearby hydrogen-bonding groups on the macromolecule, and a water molecule (H53) was in fact detected by X-ray crystallography near the center of the favored zone (Figure 2). Furthermore, the map was recontoured at more negative energy levels in order to assess the accuracy of the predictions, and this showed that the experimental position of the water molecule lies within 0.3 Å of the peak in the computed contour map.

The good agreement between observed and predicted water positions in Figure 2 might have been expected, because this particular site lies within the structure of a single phospholipase-A2 molecule and because the binding pocket is well defined by steric constraints. However, other water molecules are found close to the outer surface of the protein, where the pocket is not so obvious and where the experimental positions could in principle be influenced by interactions with neighboring macromolecules in the crystal. Whole-crystal effects of this type were not taken into account during the present calculations because they

(16) Hagler, A. T.; Dauber, P.; Lifson, S. *J. Am. Chem. Soc.* **1979**, *101*, 5131.
(17) Hagler, A. T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122.
(18) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.
(19) Snir, J.; Nemenoff, R. A.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2497.
(20) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(21) Dijkstra, B. W.; Kalk, K. H.; Hol, W. G. J.; Drenth, J. *J. Mol. Biol.* **1981**, *147*, 97.

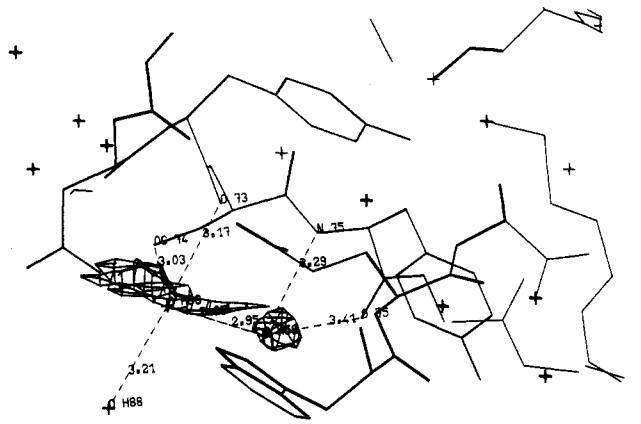
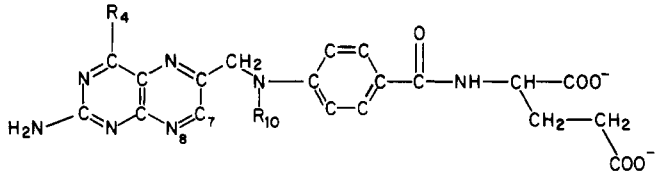



**Figure 3.** Two water molecules near the outside surface of phospholipase-A2,[21] hydrogen bonded to the protein and to each other as shown. Both lie within the −9.0 kcal/mol contours for water as expected, but another water molecule (H88) is further from the protein and is not surrounded by contours. See text.

may have little biological significance, but a typical situation near the protein surface is depicted in Figure 3. This shows a shallow pocket containing two water molecules (H59 and H96), whose observed positions are both within the calculated water contour at −9 kcal/mol as expected. However, the shape of the contours around each molecule is qualitatively different. The favored location for water H59 is clearly defined by two hydrogen bonds from N75 and O75, and its position is accordingly delineated by a globular cluster of contours peaking at the observed position. On the other hand, water H96 is surrounded by a long tube of contours showing that several different positions might be equally acceptable for this water molecule, depending on the particular hydrogen bonds formed with the protein. The actual location of water H96 appears to be influenced by the well-defined position of water H59, since H96 is observed in the contour tube at the precise place where an additional water-to-water hydrogen bond can be formed as shown in the figure.

Figures 2 and 3 demonstrate that energy contours can be of value in predicting where water molecules will bind to a protein, both inside the protein matrix and close to its surface. However, the contours become less helpful as the distance from the protein increases, and water H88 is not contained by the −9 kcal/mol water contour in Figure 3 although it is present in the experimental crystal structure. Since its distance would correspond to a second hydration layer around the protein and since its location could be influenced by intermolecular contacts within the crystal, this may not be surprising.

**Amino Contours.** The structure of *Escherichia coli* dihydrofolate reductase with the inhibitor methotrexate (I) bound at the active site has been observed by high-



I, $R_4$ = $NH_2$; $R_{10}$ = $CH_3$; aromatic
II, $R_4$ = OH; $R_{10}$ = H; 7,8-dihydro

resolution X-ray crystallography.[22] The crystals contained two conformationally distinct molecules of the enzyme, and the original authors concluded that one of these was a more

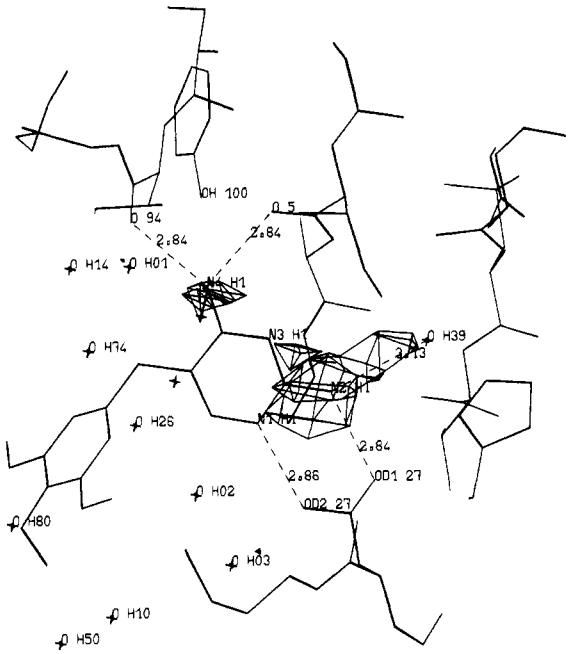
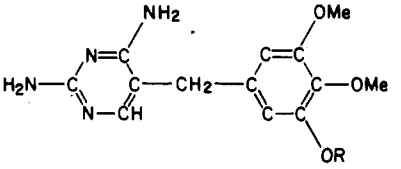



**Figure 4.** The substrate binding cleft of *E. coli* dihydrofolate reductase,[22] on which has been superimposed the enzyme inhibitor trimethoprim (molecule H1). The contours are plotted at −12.0 kcal/mol for an extended amino group. They delineate two sites of strong attraction, and the position of trimethoprim relative to the protein was adjusted so that amino nitrogens N2 and N4 of the ligand were located on these peaks. Hydrogen bonds from the ligand to the protein and to water molecule H39 are shown. See text.

representative structure than the other. Figure 4 shows the active site of the representative macromolecule, but the observed inhibitor methotrexate has been replaced in the figure by trimethoprim (III), which is another inhibitor whose structure has been independently observed.[23]



III, R = $CH_3$
IV, R = $(CH_2)_nCOOH$

The three-dimensional contours in Figure 4 were calculated for an "extended" amino group ($NH_3^+$, Table I) interacting with the protein. A GRID spacing of 1 Å was used in the calculations, since this reduces the computations by a factor of 8 without seriously degrading the quality of the contour surfaces. In fact, the contours in Figure 4 show two distinct peaks that could be of value when designing a novel drug to fit dihydrofolate reductase, and the precise location for the heterocyclic ring of trimethoprim in the figure was adjusted by placing N2 and N4 on the "extended" amino density peaks, while keeping the plane of the ring in the carboxy plane of side-chain Asp-27.

Bolin et al.[22] comment that the corresponding heterocyclic ring of methotrexate interacts with Oδ1 and Oδ2 of Asp-27 and with the carbonyl oxygens of Ile-5 and Ile-94 (*E. coli* numbering). The equivalent trimethoprim interactions are shown in Figure 4 together with a water interaction at N2, whose importance is emphasized by Bolin

(22) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.

(23) Koetzle, T. F.; Williams, G. J. B. *J. Am. Chem. Soc.* **1976**, *98*, 2074.

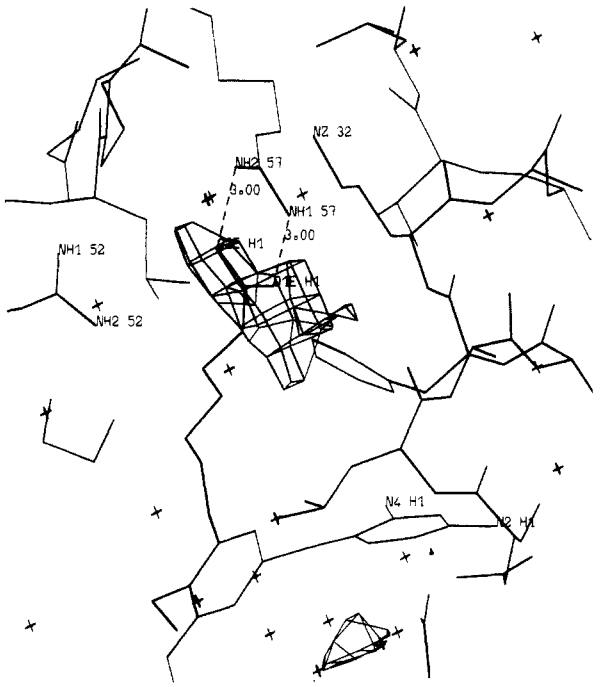



**Figure 5.** Another view of the binding cleft in dihydrofolate reductase,[22] but with contours at −4.5 kcal/mol for a carboxy oxygen atom and with a trimethoprim analogue that has a carboxy side chain. The contours indicate where the binding of this carboxy group might be expected to occur. See text. An alternative carboxy-binding region is delineated by the contours at the bottom edge of the figure.

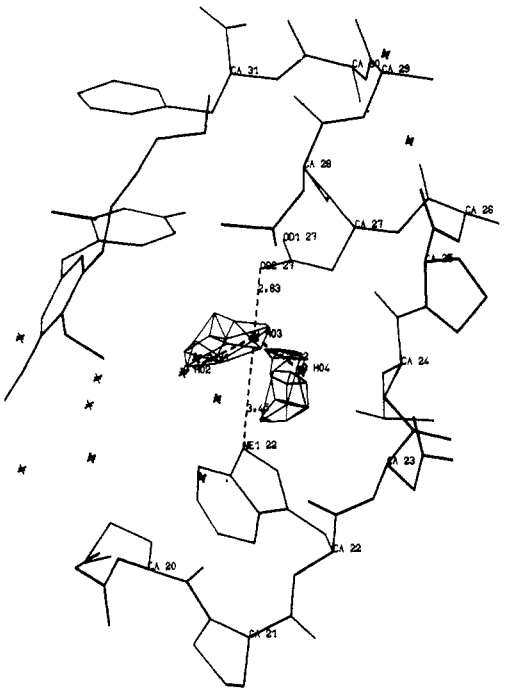



**Figure 6.** Another view of the alternative carboxy-binding region shown at the bottom of Figure 5. The contours are at −4.5 kcal/mol and surround structurally conserved water molecules.[22] Nearby residues are also well conserved, and there may be a functionally-important shuttle between Trp-22 and Asp-27. See text.

et al.[22] (H39 in Figure 4 is Bolin's Wat-639, ec2). The hydroxyl of Tyr-100 is also seen between residues Ile-5 and Ile-94 and this could have some effect on the generation of "extended" amino contours around N4. However, Tyr-100 is not a conserved residue in other dihydrofolate reductase sequences, and its hydroxyl group cannot be of general importance for the binding of diamino heterocycles to these enzymes.

**Carboxy Contours.** An analogue (IV) of trimethoprim having an $O(CH_2)_5COOH$ side chain instead of $OCH_3$ was studied by Kuyper et al.[24] Those authors showed by X-ray crystallography that the carboxy group approaches residue Arg-57 of *E. coli* DHFR, and a GRID computation has now been performed in order to find out whether this binding mode could have been predicted by using energy contours. The GRID calculation was performed with carboxy oxygen ($O^-$; Table I) as the probe atom, and the contours of Figure 5 were plotted at an energy level of −4.5 kcal/mol.

These contours were generated for a probe that simulates a single carboxy oxygen atom, but of course carboxy atoms occur as pairs in chemical groups. A small globular peak in the contour map (such as that shown in Figure 2) would not be able to accommodate a complete carboxy function. On the contrary, an elongated volume would be appropriate, and the largest contour feature on the carboxy map (Figure 5) has an elongated outline. Moreover it lies adjacent to Arg-57 in the expected orientation and is in good agreement with the observations of Kuyper et al.[24]

Another significant set of contours may be observed at the bottom edge of Figure 5, and these are redrawn in the center of Figure 6 from a viewpoint that shows that there are twin peaks surrounding a group of three water molecules. Bolin et al.[22] have drawn attention to this cluster that hydrogen bonds to the conserved residues Trp-22 and Leu-24 and to the functionally important residue Asp-27. They suggest that water in this position may be responsible for binding the oxygen substituent at position 4 of the natural substrate dihydrofolate (II), and the presence of oxygen contours in this position is not incompatible with their interpretation.

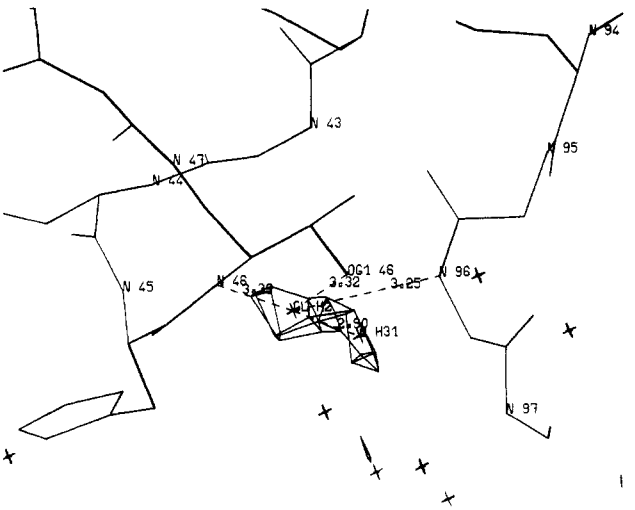



**Figure 7.** Another carboxy-binding region is indicated by these −4.5 kcal/mol contours near the amino termini of helices αC and αF of dihydrofolate reductase.[22] Such regions have been discussed by Hol et al.[26] See text. A chloride ion (Cl H2) has been observed[22] at this site, as shown.

Bolin et al.[22] have also pointed out that the α-carboxylate group in the glutamate moiety of methotrexate is bound in a well-defined pocket adjacent to residue Arg-57, but the γ-carboxylate binding is weak and nonspecific. A search was therefore made through the contour map for a site at which γ-carboxylate binding might be predicted,

(24) Kuyper, L. F.; Roth, B.; Baccanari, D. P.; Ferone, R.; Beddell, C. R.; Champness, J. N.; Stammers, D. K.; Dann, J. G.; Norrington, F. E. A.; Baker, D. J.; Goodford, P. J. *J. Med. Chem.* 1985, *28*, 303.

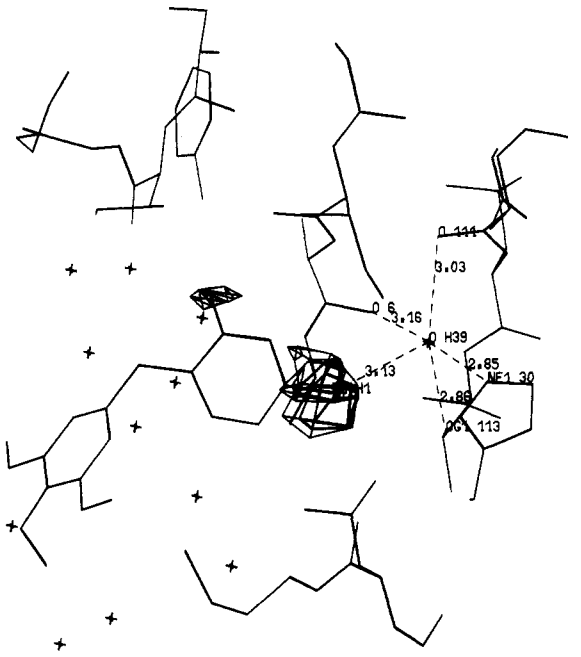



**Figure 8.** The dihydrofolate reductase binding cleft[22] with amino contours at −15.0 and (heavy lines round N2) −18.0 kcal/mol. Water molecule H39 was treated as a structural part of the protein for these calculations. The hydrogen bond from H39 to trimethoprim therefore contributes to the contours surrounding N2 of the ligand, giving a higher contour peak and a still better defined location for this atom. See text.

and a peak was actually found in the region between Oγ1 of Thr-73 and Nδ2 of Asn-59. However, these are not strictly conserved residues and neither is the nearby Lys-58, which contributes toward this peak in the carboxy contours.

A significant feature (Figure 7) was discovered on the contour map near oxygen Oγ1 of residue Thr-46. This is a conserved amino acid that hydrogen bonds directly to a phosphate oxygen of the cofactor in crystals of the ternary enzyme–methotrexate–cofactor complex from *Lactobacillus casei*.[25] The contours in Figure 7 are located near the N-terminals of α-helices C and F and apparently indicate the sort of anion binding site that has been proposed by Hol et al.[26] A chloride ion was observed here by Bolin et al,[22] in the *E. coli* DHFR crystals and is shown in Figure 7 in the zone of attraction delineated by the contours.

**Water as Part of the Protein Structure.** It may not be unreasonable to treat a water molecule that is clearly defined by a high-occupancy and low-temperature factor in the X-ray refinement, or that is indicated by a high peak in the computed energy map for a water probe, as part of the protein macromolecule itself.[27] In Figure 3, for example, water H59 might have been treated in this way, which would have altered the contours around water H96 so that its position could be more accurately predicted. Bolin et al.[22] have suggested that water H39 of the dihydrofolate reductase structure (Figure 4) comes into the same category, and the "extended" amino computations for dihydrofolate reductase were therefore repeated with H39 parameterized as a protein atom. This had the expected effect of collapsing the lobe of contour that extends in Figure 4 from N2 of trimethoprim toward H39 while sharpening the peak around N2 itself. The new contours are plotted in Figure 8 at −15.0 kcal/mol, and a second contour level is drawn with thicker lines at −18.0 kcal/mol. This higher level only surrounds the GRID sampling point at (14.0, 66.0, 47.0), which is closest to the trimethoprim N2 peak (Figure 8) and has a maximum value of −23.5 kcal/mol.

The major interactions of the "extended" amino probe at (14.0, 66.0, 47.0) will now be considered in some detail, because they illustrate the relative magnitudes of the pairwise components that contribute to the total energy value of a typical GRID point. The probe is cationic ($NH_3^+$; Table I) and the largest individual component is the electrostatic attraction $E_{el}$ of negatively charged oxygen Asp-27 Oδ1. This is 3.08 Å distant from the probe and provides −5.7 kcal/mol, while other appreciable electrostatic attractions come from Asp-27 Oδ2 (3.99 Å; −3.3 kcal/mol), Asp-27 O (3.85 Å; −3.3 kcal/mol), Ala-6 O (4.66 Å; −2.3 kcal/mol), and Trp-30 O (5.21 Å; −1.9 kcal/mol), but against these must be set electrostatic repulsions from Asp-27 Cγ (3.90 Å; +2.7 kcal/mol), Trp-30 C (4.75 Å; +2.2 kcal/mol), and Asp-27 C (4.88 Å; +2.1 kcal/mol). These are significant energies, and atoms as far away as 6 Å can still make an individual electrostatic contribution exceeding 1 kcal/mol, while the total electrostatic term for all protein atoms more distant than this is a net attraction of −4 kcal/mol. Clearly one cannot neglect the electrostatic interactions of distant atoms when a charged probe is used, and the treatment of the dielectric effect is therefore important.[14]

The hydrogen bonds from water H39 and Asp-27 Oδ1 each have a computed strength of −3.0 kcal/mol, and there are some 15 protein atoms within 4 Å of probe position (14.0, 66.0, 47.0), each contributing over −0.2 kcal/mol to the Lennard–Jones attraction. Only Trp-30 Cδ1 generates a net Lennard–Jones repulsion, and this is smaller (+0.1 kcal/mol) than the electrostatic repulsion (+0.2 kcal/mol) due to the positive charge on the same protein atom. Hence the hydrogen bonds and Lennard–Jones forces make up roughly half the interaction, while electrostatic effects account for the remainder. However, these proportions depend on the choice of probe group and the parameters that describe it.

Another GRID computation was performed in which water molecules HO3 and HO4 (Figure 6) were parameterized as a part of the *E. coli* dihydrofolate reductase structure. Contours were calculated with carbonyl oxygen (O; Table I) as the probe group, in order to follow the suggestion of Bolin et al.[22] that this cluster of waters might facilitate the binding of the natural substrate dihydrofolate. However, no evidence was obtained to support their interesting hypothesis because no contour peak was observed in the expected location. The highest carbonyl peak in the *E. coli* dihydrofolate reductase binding cleft was actually near N4 of trimethoprim (Figure 4), but the hydrogen bond from Tyr-100 made a significant energy contribution to this value and Tyr-100 is not a conserved residue.

The presence of a significant contour peak near N4, and its absence near water molecules HO3 and HO4 do not necessarily run counter to the proposals of Bolin et al.[22] The oxygen atoms at position 4 of dihydrofolate might replace a bound water molecule instead of binding to it, or the wrong water molecules might have been selected for inclusion as part of the protein structure in the present calculations, or the water positions might alter when dihydrofolate reductase interacts with substrate, or the

(25) Filman, D. J.; Bolin, J. T.; Matthews, D. A.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13663.
(26) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. *Nature (London)* **1978**, *273*, 443.
(27) Edsall, J. T.; McKenzie, H. A. *Adv. Biophys.* **1983**, *16*, 53.

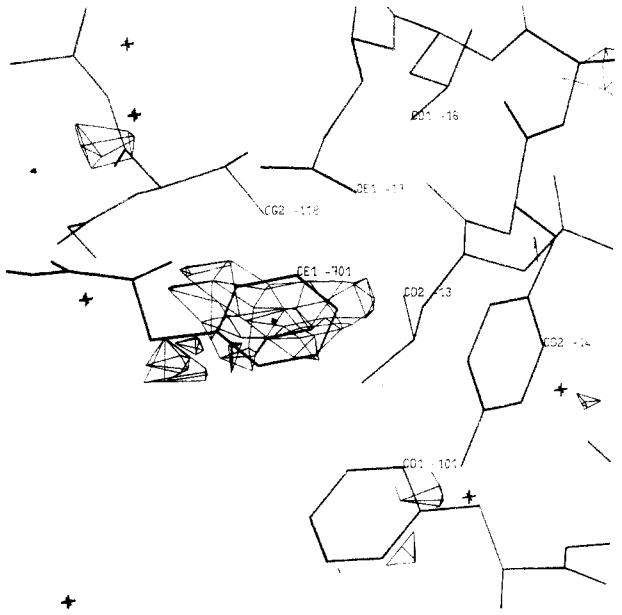


**Figure 9.** Residues 13, 14, 16, 17, 101, and 118 are all part of one insulin monomer whose structure[28] was used to calculate contours for an extended methyl group. These contours delineate a cleft that is filled by the side chain of residue Phe-701, coming from a neighboring insulin monomer in the crystal, thus demonstrating how the contours can be used to predict protein–protein interactions. See text.

parameterization of the probe as a carbonyl group might have been an inappropriate representation for the oxygen substituent at position 4 of the substrate dihydrofolate. However, the use of hydroxy (OH; Table I) as an alternative probe group did not provide any more support for the hypothesis of Bolin et al.[22]

**Protein–Protein Interactions.** The present model simulates Lennard–Jones, electrostatic, and hydrogen-bond interactions, and these same effects should influence other systems besides the binding of small ligands to proteins. An attempt was therefore made to use energy contours in order to investigate protein–protein interactions.

Insulin consists of two different protein chains joined together by disulfide bridges to form a single monomeric subunit. GRID computations were performed with an "extended" methyl probe ($CH_3$; Table I) on the porcine insulin monomer,[28] and the strongest peak on the map is shown in Figure 9. The parameters for "extended" methyl were assigned on the assumption that it does not interact electrostatically with the protein and that it does not form hydrogen bonds. This only leaves the Lennard–Jones interaction, which is relatively weak, and the contours in Figure 9 are accordingly plotted at −1.5 kcal/mol, which corresponds to a rather modest attraction.

Residues 13, 14, 16, and 17 are in the first protein chain, and residues 101 and 118 are in the second chain of the same insulin subunit. All these residues therefore contribute toward the contours in Figure 9, but insulin monomers can associate into dimers and hexamers, and residue Phe-701 does not come from the same monomer but from a neighboring monomeric subunit in the crystal. The zone of attraction from the first monomer envelopes the hydrophobic groups of this Phe-701 side chain from the second as shown in the figure. In this case the contours seem to delineate an hydrophobic pocket which is exploited to facilitate a monomer–monomer association in the hexameric crystalline structure, thus demonstrating how such contours can indicate significant features of protein–protein interactions.

## Discussion

Before assessing the findings, it is necessary to consider the shortcomings of the method. The positions of the protein atoms were observed by X-ray crystallography, and the strengths and weaknesses of that technique have recently been reviewed in the context of drug design.[1] The role of entropy is a very significant factor in ligand binding since the release of solvent molecules exerts an important influence in the hydrophobic effect,[29] but entropy was neglected when computing the present energy contours. Moreover, the decision to divide up the ligand into separate probe groups and the assumption of pairwise additivity between atoms are at best unjustifiable approximations. Furthermore an "extended" atom may have one set of characteristics in proteins but behave differently in other chemical environments. Thus the amino group may be $NH_3^+$ in the side chain of lysine and has been treated as such by Brooks et al.[11] and in the present work, but the amino substituents at positions 2 and 4 of trimethoprim may be better represented by $NH_2$. The fit of trimethoprim to the "extended" $NH_3^+$ contours is therefore surprising at first sight and may indicate that more of the positive charge on the cationic form of the bound heterocycle is delocalized to the amino substituents than is currently thought.

This raises another problem because the distribution of electric charge may be altered in the heterocyclic and aromatic rings of trimethoprim when it interacts with the protein, and such alterations are in no way modelled by the "extended" amino group of Figure 4. The reality of these effects is clearly illustrated by changes in the $pK_a$ of ionizable groups in the ligand when it binds to the macromolecule,[30] but they can also occur in systems where they are not detected so easily, such as the nearby amino acid residue Phe-31 of dihydrofolate reductase.

The local conformation of the macromolecule in the observed X-ray structure may be molded around any bound ligand in the crystal. Indeed, this is probably a partial explanation for the good fit of ligands that has been observed by X-ray crystallography. It has been shown that 2,3-diphosphoglycerate influences the conformation of hemoglobin[31] and that the structure of cytochrome *c* changes on oxidation.[32,33] Such movements of atoms and charges in ligand and macromolecule from their initially preferred positions could be energetically expensive, and an appropriate part of this deformation energy should perhaps be set against $E_{xyz}$. It was not offset in the present calculations, and the resulting GRID values may well be biased.

The above shortcomings are not directed against energy contours in particular. Similar problems arise with conventional space-filling models, and they do not yield so much useful information because they only define forbidden conformations. The present approach allows the detailed attributes of an energy profile to be studied in relation to the structure of a protein, and this can be particularly helpful when novel ligands are being designed to fit a macromolecule. The most favorable sites for each

(28) Isaacs, N. W.; Agarwal, R. C. *Acta. Crystallogr., Sect. A* **1978**, *A34*, 782.
(29) Tanford, C. "The Hydrophobic Effect: Formation of Micelles and Biological Membranes", 2nd ed.; Wiley: New York, 1980.
(30) Roberts, G. C. K.; Feeney, J.; Burgen, A. S. V.; Daluge, S. *FEBS Lett.* **1981**, *131*, 85.
(31) Arnone, A. *Nature (London)* **1972**, *237*, 146.
(32) Warshall, A.; Churg, A. K. *J. Mol. Biol.* **1983**, *168*, 693.
(33) Churg, A. K.; Weiss, R. M.; Warshel, A.; Takani, T. *J. Phys. Chem.* **1983**, *87*, 1683.

kind of probe are displayed, together with the energetic penalty that must be paid if the probe group is not in exactly the best place. Moreover, the closeness of the contours indicates what forces the probe would be experiencing and their direction.

The usefulness of energy contours might be diminished if there were many spurious peaks which could mislead the observer. However the present results show that the major features of a predicted interaction can be studied while smaller peaks are automatically filtered out by the contouring procedure. The significance of these major effects may be judged from the figures, but the amount of information on view is easily varied by recontouring at different energy levels. This recontouring is a quick process that does not require another run of programs GRIN and GRID. On the contrary, it can be done while the energy maps are being studied, so that the user directly controls the amount of information that he is considering at any time.

The ability of different probes to interact favorably with proteins appears to follow a general ranking order. It is not uncommon to find sites where the "extended" amino group has a computed energy of −15 kcal/mol, and values as high as −23 kcal/mol have been observed. At the other end of the scale comes the "extended" methyne group (CH), which has a low polarizability and does not hydrogen bond at all. "Extended" methyl ($CH_3$) is a little better but rarely achieves more than −4 kcal/mol, while carbonyl oxygen may reach −7 and carboxy oxygen −10 kcal/mol. One cannot be precise about the ranking order, and it must be borne in mind that a complete carboxy group contains two oxygen atoms while the computation relates to one alone. Although no quantitative claims should be made for the actual contour levels, these observations suggest that amino and carboxy functions might be used with advantage when high-affinity ligands are being designed.

The present examples mostly come from well-known binding sites on well-known proteins, but contour peaks often indicate favored sites where no ligand other than water has been observed so far. For example, a water molecule has been reported in the pig insulin structure at hydrogen bonding distance from a backbone nitrogen and a backbone carbonyl and Glu-4 Oε1. The highest peak on the contour map for an "extended" amino probe occurs in precisely the same location with a computed energy of −18 kcal/mol, making this the most favored site–probe interaction anywhere on the porcine insulin molecule.

Sites like this, or the hydrophobic pocket illustrate in Figure 9, may be places where one macromolecule makes biologically significant contacts with another in vivo. On the other hand, nature may not use all the favored sites that occur on proteins, but they could still be appropriate targets for novel therapeutic agents. The α-terminal site of hemoglobin[34] may be such an example, and although the choice of energy functions at any one time is always open to improvement, the present contour maps suggest that there really are favored sites on macromolecules and show that energy contours may provide an effective approach to the design of high-affinity ligands as drugs.

**Acknowledgment.** I thank Professor Sir David Phillips for his support and encouragement. I am greatly indebted to him and to Margaret Adams, Peter Artymiuk, Jane Burridge, Will Gibson, Sheila Gover, Richard Pickersgill, Graham Richards, Neil Rogers, Michael Sternberg, and Garry Taylor for all their suggestions, advice, and help.

**Registry No.** I, 59-05-2; III, 738-70-5; IV ($n$ = 5), 82830-33-9; $H_2O$, 7732-18-5; phospholipase-A2, 9001-84-7; dihydrofolate reductase, 9002-03-3; porcine insulin, 12584-58-6.

(34) Beddell, C. R.; Goodford, P. J.; Kneen, G.; White, R. D.; Wilkinson, S.; Wooton, R. *Br. J. Pharmacol.* **1984**, *82*, 397.

# Comparative Computer Graphics and Solution Studies of the DNA Interaction of Substituted Anthraquinones Based on Doxorubicin and Mitoxantrone

Suhail A. Islam,[†] Stephen Neidle,[†] Bijukumar M. Gandecha,[§] Malcolm Partridge,[‡] Laurence H. Patterson,[‡] and Jeffrey R. Brown*[§]

*Cancer Research Campaign Biomolecular Structure Research Group, Department of Biophysics, King's College, London WC2B 5RL, School of Pharmacy, Leicester Polytechnic, Leicester LE1 9BH, and Department of Pharmaceutical Chemistry, Sunderland Polytechnic, Sunderland SR2 7EE, England.* *Received July 19, 1984*

1-[[(Diethylamino)ethyl]amino]- and 1,4-, 1,5-, and 1,8-bis[[(diethylamino)ethyl]amino]anthraquinones are shown to intercalate into DNA. Computer graphics modelling of their intercalation into the self-complementary deoxydinucleoside d(CpG) showed differences in binding properties. While the 1-substituted compound can bind from either groove, the 1,8-disubstituted compound binds with both substituents in the major groove. In the low-energy state of the complex with the 1,5-disubstituted compound, this ligand "straddles" the site with a substituent in each groove—to do this, the compound must bind to a non-base-paired region, so inducing base pairing. The 1,4-compound binds from the major groove; "straddling" is also possible if full minimization of deoxydinucleoside geometry is performed. The differences in binding mode and interaction energies are reflected in the affinities of interaction (1,5- > 1,4- ≫ 1,8- > 1-); also the antiproliferative effects in vitro are in general agreement with this ranking.

Doxorubicin (1) has a wide spectrum of antitumour activity,[1] and although it shows other cellular effects,[2,3] its cytotoxic effect is probably a result of its intercalative interaction with nuclear DNA.[4] This DNA interaction could lead to the cytotoxicity either directly (by inhibition of DNA and RNA synthesis)[5] or indirectly (by causing single or double strand breaks).[6] In an attempt to mimic

[†] University of London.
[‡] Leicester Polytechnic.
[§] Sunderland Polytechnic.

(1) Young, R. C.; Ozols, R. F.; Myers, C. E. *N. Engl. J. Med.* **1976**, *305*, 139.
(2) Moore, H. W.; Czerniak, R. *Med. Chem. Rev.* **1981**, *1*, 249.
(3) Tritton, T. R.; Yee, G.; Wingard, L. B. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **1983**, *42*, 284.
(4) Brown, J. R. *Prog. Med. Chem.* **1978**, *15*, 125.
(5) DiMarco, A. *Antibiot. Chemother.* **1978**, *23*, 216.

0022-2623/85/1828-0857$01.50/0 