Representation of noncovalent interactions in protein structures

Xavier de la Cruz, Jesús Reverter and Ignacio Fita

Departamento de Ingenieria Quimica, Escola Tècnica d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Barcelona, Spain

The energetics of solvent–atom and atom–atom nonbonded interactions can be described, for protein structures, in terms of the accessible[1] and the contact[2] atomic surface areas, respectively. This type of description emphasizes the importance of the local environment around groups in the three-dimensional structure of protein molecules. The graphical representation of nonbonded interactions according to this description allows one to visualize the spatial extent and distribution of these interactions and the relative stability of atoms or atomic groups in known or modified protein conformations. Applications of this short range description and of its graphical representation will be discussed.

Keywords: noncovalent interactions in protein structures, stability of protein conformations, free energy of solvation

INTRODUCTION

The exact nature of noncovalent interactions, including solvent interactions, appears to have a decisive importance in determining the structural stability and the specific recognition mechanisms in biomacromolecules.[3] For the empirical or semiempirical potentials, often used to model the intensity of atom-atom non-bonded interactions, an infinite distance range should, in principle, be considered.[4,5] However statistical analysis of residue–residue distances in protein structures[6] indicate that the effective influence of the contact becomes negligible for interresidue distances beyond about 10 Å. In fact, shorter cutoff distance values (6–8 Å) are often used for atom–atom nonbonded interactions in molecular mechanics and dynamical simulations of macromolecules both to decrease the computational burden and to improve the behavior of the system during the molecular simulation.[7–9] Also, a short-range representation[10,11] has been proposed to evaluate the interactions of a molecule with its surrounding solvent in which only the solute atoms or atomic groups directly involved in the interaction are considered. In this representation the intensity (ΔG_i^S) of the interaction of solvent with a functional group i of the solute is directly proportional to the solvent-accessible surface S_i^0 of group i,[1,12] the total free energy (ΔG^S) of solvation of a solute molecule is then expressed as:

$$\Delta G^S = \sum_i \Delta G_i^S = \sum_i g_i S_i^0 \quad (1)$$

where the summation extends over all the groups in the solute, and the constant of proportionality g_i represents the contribution to the free energy of hydration of group i per unit of accessible area. Thus, it appears, that a short-range description of noncovalent bonds that takes into account both the molecule–solvent and the intramolecular interactions should reflect the dominant contributions of these interactions in protein structures. This information, being spatially localized, is well suited for a diversity of graphical displays and subsequent analysis.

In this work we present a parametrization of the atom–atom noncovalent interactions based on the accessible and contact atomic areas in three-dimensional models of protein structures. This parametrization, when solvent–atom interactions also are considered, provides a formally unified description of all the noncovalent interactions acting in a protein conformation. This description has a short-range character as only surfaces spatially in proximity in the three-dimensional molecular structure can make a significant contribution. Finally, graphical examples displaying peculiarities of local environments in protein structures will be shown.

METHODOLOGY

The contact area (S_{ij}) between a source atom i and a target atom j can be defined as the amount of the surface from atom i that becomes inaccessible when the contacting atom j is considered.[13] The intensity of the nonbonded interactions (ΔG_{ij}) between those atoms could be linearly related, when there is no short contact between them, with the extent of their contact areas. Thus,

$$\Delta G_{ij} = \lambda_{ij} S_{ij} \quad (2)$$

where λ_{ij} would be the linear coefficient parametrizing the interaction per unit of contact area between atoms i and j. In this work the λ_{ij} values (Table 1) have been obtained by adjusting Equation (2) to the empirical potentials used in the program package AMBER[14] (Figures 1, 2). With the

Color Plates for this article are on page 110.

Address reprint requests to Dr. Fita at the Departamento de Ingenieria Quimica, Escola Tècnica d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Avinguda Diagonal, 647, 08028 Barcelona, Spain.
Received 28 March 1991; accepted 7 October 1991

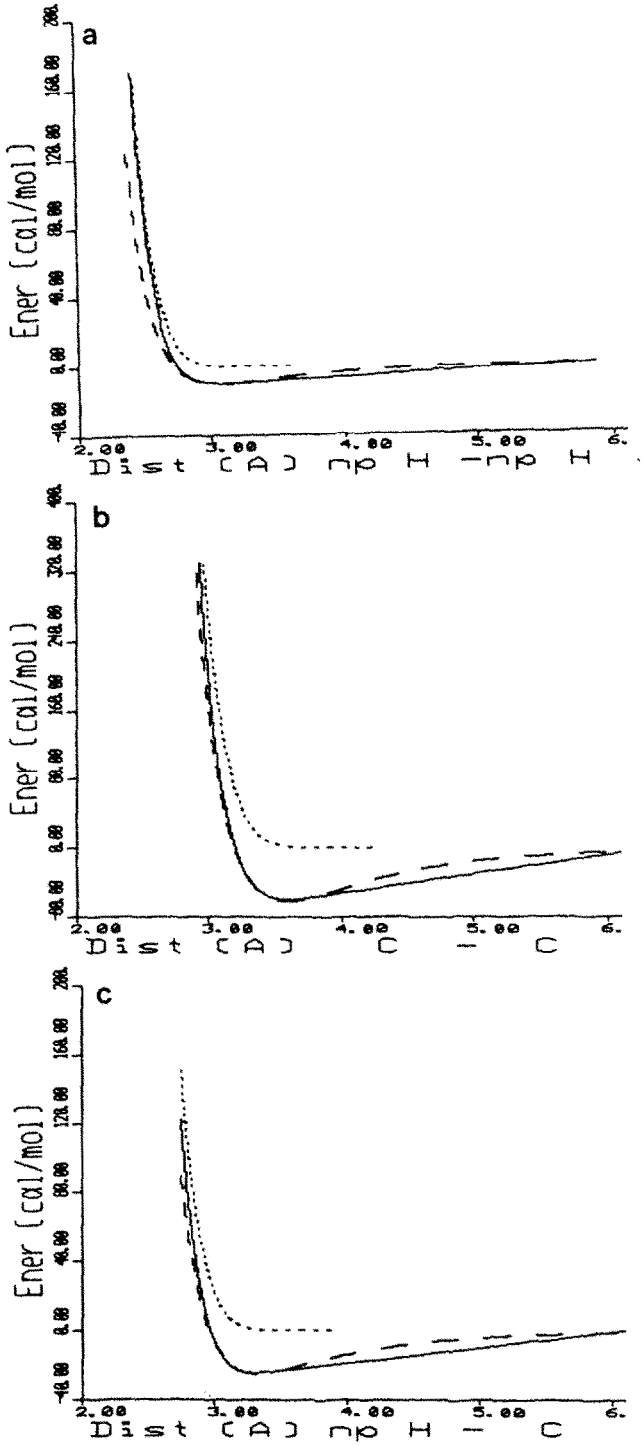

Figure 1. Nonbonded potentials for the noncovalent interactions (a) Apolar H-Apolar H, (b) $C(sp^3)$-$C(sp^3)$ and (c) H-$C(sp^3)$. Potentials derived from the AMBER parameters without the Coulombic contributions are indicated as a discontinuous line.[14] The "empirical" energies, computed using Equation (2) in the text, are drawn with a continuous line. The repulsive component due to the excluded volume effects of short contacts is indicated by dots. The quality of the adjustment between the AMBER potentials and these empirical energies is better than the accord between different sets of potentials (de la Cruz and Fita, in preparation).

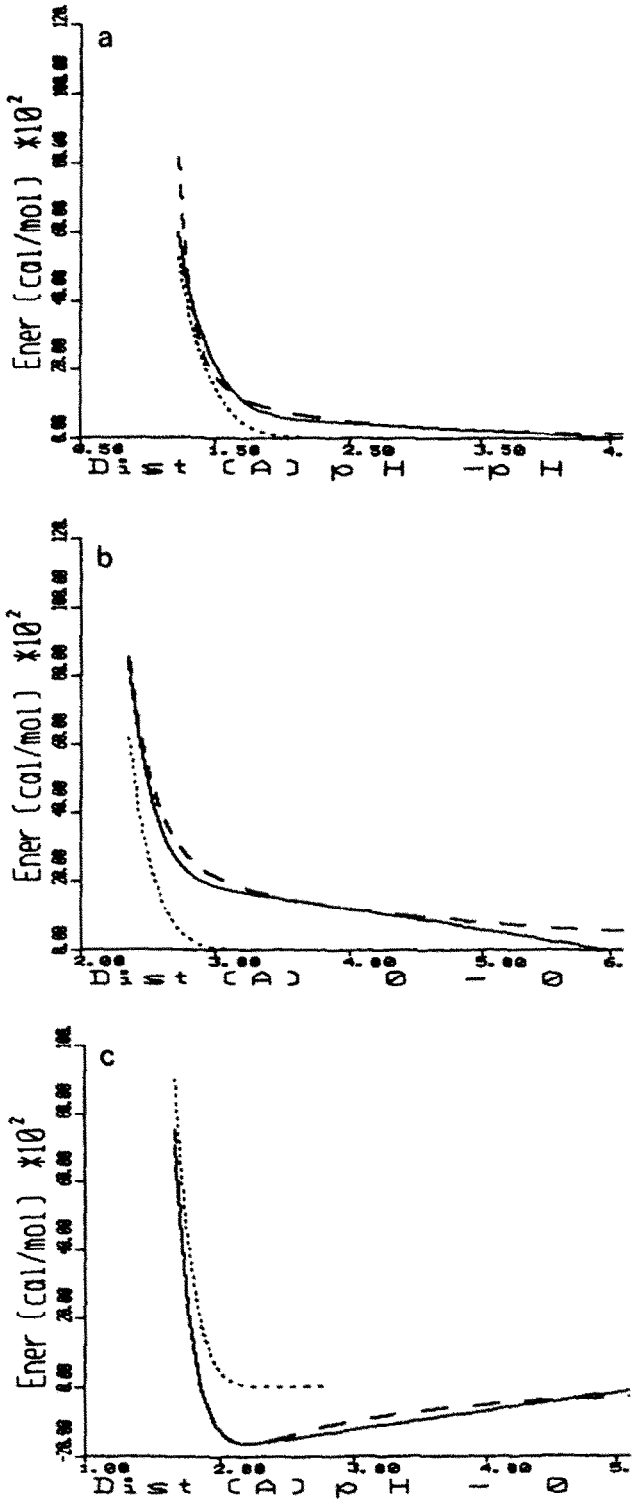

Figure 2. Nonbonded potentials for the noncovalent interactions (a) Polar H-Polar H, (b) Polar O-Polar O and (c) Polar H-Polar O. Potentials derived from the AMBER parameters[14] with the Coulombic contribution included are indicated as a discontinuous line. A dielectric constant of 4R as suggested by Whitlow and Teeter[7] was used. The "empirical" energies computed using the interacting surfaces (see text) are drawn with a continuous line; the repulsive component due to the excluded volume effects of short contacts is indicated by dots.

assumption that noncovalent interactions have additive contributions,[15] the nonbonded "empirical" energy of atom i (ΔG_i^{NB}) due to all its atom–atom interactions can be expressed as:

$$\Delta G_i^{NB} = \sum_{j=1} \Delta G_{ij} = \sum_{j=1} \lambda_{ij} S_{ij} \quad (3)$$

If atom i is also interacting with the solvent, the solvation free energy can be added to Equation (3), using expression Equation (1), giving

$$\Delta G_i^{E} = \Delta G_i^{S} + \Delta G_i^{NB} = \sum_{j=0} \lambda_{ij} S_{ij} \quad (4)$$

where the summation extends over all the interacting surfaces (S_{ij}) that the atom i has. For $j = 0$ the corresponding interacting surface S_{i0} is the accessible surface of atom i (S_i^0) and λ_{i0} coincides with the hydration coefficient g_i. Thus for atom i, ΔG_i^E describes in terms of empirical energy the intensity of both atom–atom and atom–solvent noncovalent interactions. The total empirical energy of a protein molecule, due to these noncovalent interactions, would then be:

$$\Delta G = \sum_i \Delta G_i^{E} = \sum_i \sum_{j=0} \lambda_{ij} S_{ij} \quad (5)$$

where the summation on i includes all protein atoms. Again, for clarity possible contributions due to the presence of short contacts in the structure have been omitted.

The description proposed allows one to treat all the noncovalent interactions in a protein structure as being formally equivalent. This unified description can be used to evaluate the local relative stabilities of protein conformations in different situations. In this work two cases have been considered: the stability of folded vs. unfolded conformations; and the use of a reference ideal state in order to optimize a protein structure.

The stability of a folded conformation ($\Delta\Delta G^{\alpha}$) is usually described[3] as the difference in free energy between the unfolded ($\Delta G_{unfolded}$) and folded (ΔG_{folded}) free energies under certain environmental conditions:

$$\Delta\Delta G^{\alpha} = \Delta G_{folded} - \Delta G_{unfolded} = \sum_i (\Delta G_{i,\ folded} - \Delta G_{i,\ unfolded}) = \sum_i \Delta\Delta G_i^{\alpha}$$

Thus $\Delta\Delta G_i^{\alpha}$ corresponds to the stability of atom or atomic group i in the conformation.

Assuming that in the unfolded state the molecule is fully accessible to the solvent, which in most cases is only an approximation, we can use Equation (4) to evaluate $\Delta\Delta G_i^{\alpha}$ as

$$\Delta\Delta G_i^{\alpha} \approx \sum_{j=1} (\lambda_{ij} - \lambda_{i0}) S_{ij} = \sum_{j=1} \alpha_{ij} S_{ij} \quad (6)$$

The α_{ij} coefficients can be positive, meaning that the solvation is more favorable than the ij interaction, or can be negative in the opposite case.

To evaluate the stability of an interaction for a given source atom i, the interaction could be compared with the most stable interaction (the ideal interaction, λ_{il}) the atom can make. A global deviation from ideality ($\Delta\Delta G^{\beta}$) of the noncovalent interactions acting in a protein structure can be expressed as:

$$\Delta\Delta G^{\beta} = \Delta G - \Delta G^{ideal} = \sum_i \sum_{j=0} (\lambda_{ij} - \lambda_{il}) S_{ij} = \sum_i \sum_{j=0} \beta_{ij} S_{ij} \quad (7)$$

where ΔG^{ideal} corresponds to an "ideal" conformation in which every atom i would make only its most favorable interactions. All the β_{ij} coefficients are, by construction, positive. They become 0 when the observed interaction ij corresponds to the best interaction of atom i.

Minimization of Equation (7) gives the closest configuration to the ideal situation, and thus the configuration with the lowest empirical energy. As every term in Equation (7) is positive or null, the minimization of this expression can be done using a least-squares algorithm. We have added[2] Equation (7) to the well-known Hendrickson and Konnert refinement package[16] as a new observational function (ϕ_{int}) for the noncovalent interactions. In the least squares formalism ϕ_{int} can be expressed as:

$$\phi_{int} = \sum_m^{inter} 1/\sigma^2(m)\ (I_m^{ideal} - I_m^{model})^2 \quad (8)$$

where m corresponds to the jth interaction of atom i. Taking $\sigma^2(m) = 1/\beta_{ij}$ and $I_m^{model} = \sqrt{S_{ij}}$, we recover Equation (7), since I_m^{ideal} would vanish for all possible interactions but the best, in which case $\beta_{ij} = 0$.

RESULTS AND GRAPHICAL REPRESENTATIONS

The structure of native BPTI (Protein Data Bank reference pdb4pti), with all the hydrogen atoms added in positions optimized stereochemically and for their nonbonded interactions,[17] and was analyzed according to the methodology presented above.

The relative stabilities of residues along the chain (Figure 3) were calculated using Equation (6) for the atom types whose λ_{ij} coefficients are given in Table 1. These relative stabilities indicate that even with the reduced set of coef-

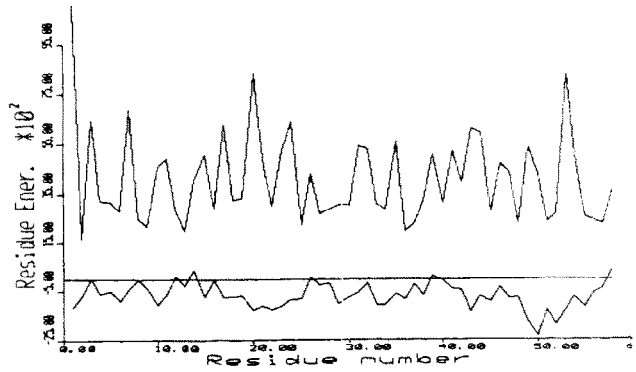

Figure 3. Stability energies of residues in BPTI. The values for each residue were obtained adding the corresponding values of every atom in the residue. The atomic values in turn were computed using Equations (6) and (7) in the text. The reduced set of λ_{ij} values used in these calculations were those in (Table 1) and de la Cruz and Fita.[2]

Table 1. Derived coefficients[a] from the potentials[b]

	Nonpolar H	Polar H	C	O	N	S
Nonpolar H	−2					
Polar H	−2	77				
C	−4	−5	−11			
O	−8	−211	−19	286		
N	−7	−151	−18	212	156	
S	−8	−38	−21	22	11	−27

[a] Different sets of atomic surfaces could be defined according to the type of structural analysis
[b] The Amber potentials[14] were used. The dielectric constant was taken as $4R$ as recommended by Whitlow and Teeter[7]

ficients used only five residues, GLY 12, CYS 14, LYS 26, ARG 39 and ALA 58, appear to be marginally unstable in the native conformation. The relative idealities were calculated using Equation (7), again for the atom types in Table 1. As expected, there is not a single residue having fully ideal interactions, though all the residues, with the exception of the five indicated above, are stable.

Using Equation (5) the intensity of interactions between atoms in aromatic residues was quantified for the BPTI structure. The results are graphically represented (Color Plate 1) with a density map in which the height of peaks associated with atoms has been taken proportional to the extent of the contacts between atoms in those aromatic side chains. In BPTI only five aromatic residues, PHE 4, PHE 22, PHE 33, TYR 35 and PHE 45, participate in these aromatic–aromatic interactions. Residue PHE 33 has the strongest interactions of this kind, both for the number of atoms involved and for the extent of the contacts. Residue TYR 35 makes a few, but quite strong, aromatic contacts.

Noncovalent interactions of polar hydrogen atoms were examined for the BPTI structure using Equation (5). Again a graphical representation using density maps, as described before, was done (Color Plate 2). The heights of the peaks were taken to be proportional to the surface area of polar hydrogen atoms that was not accessible to the solvent nor in contact with polar oxygens. Three strong peaks were detected. The most intense spot was around the terminal amino groups of residues ASN 24 and GLN 31. The second spot was the terminal polar hydrogen atom in the side chain of TYR 35. A closer analysis of these two poor environments suggested, in both cases, a modification of the atomic BPTI model used. Thus it appears that the terminal group of GLN 31 should be rotated, interchanging the positions of nitrogen and oxygen in the side chain of this residue. Also problems for residue TYR 35 disappear by simply shifting the dihedral angle of the terminal polar hydrogen by 180°. For this hydrogen the automatic optimization used to refine the positions of the hydrogen atoms added to the Protein Data Bank BPTI coordinates had not been accomplished.

All the calculations presented in this work were done using an IRIS (Silicon Graphics) 4D/70G workstation. For BPTI each cycle of refinement had a duration of about 60 seconds. The most time-consuming step, the contact and accessible surface areas calculation, was done[2] using a modified version of Richard's algorithm.[1]

CONCLUSIONS

The quantitative description proposed in this work[2] offers a unified treatment of the nonbonded interactions acting in a given protein conformation; thus the energetics of atom–atom noncovalent interactions and of solvation can be analyzed and represented similarly. This description easily allows one to evaluate peculiarities and changes in the local environment of atomic groups in different protein conformations. Therefore it appears well suited to check the quality of molecular models and the potential reactivity of different parts of the molecule structure. The graphical display is well suited to the local character of this approach.

ACKNOWLEDGEMENTS

We would like to thank J. Ferrán and J.M. Zaldivar for their suggestions, and J.A. Subirana for his constant support. This work was made possible in part by grants BT87-0009 and BIO756/90 from the Plan Nacional de Biotecnología CICYT (Spain); X. de la Cruz acknowledges a predoctoral fellowship from the Departament d'Ensenyament de la Generalitat de Catalunya.

REFERENCES

1 Lee, B. and Richards, F.M. *J. Mol. Biol.* 1971, **55**, 379–400
2 de la Cruz, X. and Fita, I. *J. Appl. Crys.* 1991, **C47**, 941–946
3 Dill, K.A. *Biochemistry* 1990, **29**, 7133–7155
4 Rogers, N.K. *Prediction of Protein Structure and the Principles of Protein Conformation.* (G.D. Fassman, Ed.) Plenum Press, New York (1989) 359–389
5 Burley, S.K. and Petsko, G.A. *Advances in Protein Chemistry* (C.B. Anfinsen, J.T. Edsall, F.M. Richards and D.S. Eisenberg, Eds.) Academic Press, San Diego (1988) vol. 39, 125–189
6 Viswanadhan, V.N. *Int. J. Biol. Macromol.* (1987) **9**, 39–48
7 Whitlow, M. and Teeter, M.M. *J. Am. Chem. Soc.* (1986) **108**, 7263–7172
8 Brooks, C.L., Pettitt, B.M. and Karplus, M. *J. Chem. Phys.* (1985) **83**, 5897–5908
9 Mackay, D.H.J., Cross, A.J. and Hagler, A.T. *Prediction of Protein Structures and the Principles of Pro-*

tein Conformation (G.D. Fassman, Ed.) Plenum Press, New York (1989) 317–358
10 Eisenberg, D. and McLachlan, A.D. *Nature* (London) 1986, **319**, 199–203
11 Ooi, T., Oobatake, M., Némethy, G. and Scheraga, H.A. *P.N.A.S.* 1987, **84**, 3086–3090
12 Shrake, A. and Rupley, J.A. *J. Mol. Biol.* 1973, **79**, 351–371
13 Richmond, T.J. and Richards, F.M. *J. M. Biol.* 1978, **119**, 537–555
14 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. *J. Am. Chem. Soc.* 1984, **106**, 765
15 Tanaka, S. and Scheraga, H.A. *Macromolecules* 1976, **9**, 945–950
16 Hendrickson, W.A. and Konnert, J.H. *Computing in Crystallography* (R. Diamond, S. Ramaseshan and K. Venkatesan, Eds.) Bangalore, Indian Academy of Sciences (1980) 1301–1325
17 Brunger, A.T. and Karplus, M. *Proteins* 1988, **4**, 148–156
18 Cambillau, C. and Horjales, E. *J. Mol. Graph.* 1987, **5**, 174