EVALUATION OF "FREE ENERGY" FOR NON-COVALENT INTERACTIONS IN PROTEIN STRUCTURES

X. DE LA CRUZ AND I. FITA

Departament d'Enginyeria Química. ETSEIB.
Universitat Politècnica de Catalunya

Introducció

For a large number of protein it has now been stablished that their natives structures are thermodynamically stable (Santoro & Bolen, 1988) (The thermodynamic hipothesis (Anfinsen, 1973)). Thus in these proteins the native structure corresponds, by definition, to the macroscopic state having a global minimum of th free energy relative to all other states accessible on that time scale (Dill, 1990). The free energy of a protein structure consists of the sum of contributions from its intramolecular interactions and from the interaction of the molecule with the surrounding solvent water. Exact computation of those contributions sitll poses problems specially when effects with a strong entropic component, such as the hydrophobic forces or the conformational entropy, have to be evaluated (Ooi et al., 1987). Thus the accurate modeling of protein folding processes appears still as a distant goal (Karplus & Petsko, 1990).

There are different approaches to evaluate the interactions of a molecule with its surrounding solvent (Finney & Savage, 1988; Warshel & Creighton, 1989), in particular, it has been proposed (Eisenberg & McLachlan, 1986; Ooi et al., 1987) that the extent of the interaction of any functional group i of the solute with the solvent is proportional to the solvent-accessible surface area S_i^0 (Lee & Richards, 1971; Richards, 1977) of group i. The total free energy of hydration of a solute molecule is then expressed as:

$$\Delta G^S = \Sigma_i \, g_i \, S_i^0 \qquad (1)$$

where the summation is for all the functional groups in the solute and the constant of proportionality g_i represents the contribution to the free energy ofhydration of group i per unit of accessible area.

To compute the intensity of the atom-atom non-covalent interactions, in the solute molecule, empirical or semiempirical potentials are often used (Levitt, 1982; Némethy et al., 1983; Karplus & McCammon, 1983; Weiner et al., 1986; Sippl, 1990). However it has also been shown that the extent of the atomic contact areas in protein structures allows a description, at least qualitative, of the intensity of these non-covalent interactions (Fita et al., 1986; Miller et al., 1987; Lesser & Rose, 1990). Contact area (S_{ij}) between a source and a target atom i and j respectively can be defined as the amount of surface area of the source atom that becomes inaccessible when the target atom is considwered (Richmond & Richards, 1978). The characterization of accessible and interacting surface sizes and shapes appears thus to provide important information for the analysis of all kinds of non-covalent interactions in protein structures (Janin, 1979; Rose & Lee, 1986; Janin et al., 1988; Silla et al., 1990, Rowland et al., 1990).

In this work we show how the description of atom-atom interactions using contact areas can be parametrized to reproduce accurately the empirical potentials used in packages such as AMBER (Weiner et al., 1984). Then an algorithm to minimize the structural free energy is proposed in which both solvent-molecule and intramolecular non-covalent interactions are formally treated equal. Some peculiarities of this unified description representation will be discussed.

Methodology

The "empirical" energy of atom i (ΔG_i^E) due to its non-covalent interactions with the remaining atoms in the molecule, when no short contacts are present, is expressed as:

$$\Delta G_i^E = \Sigma_{j=1} \lambda_{ij} S_{ij} \qquad (2)$$

where λ_{ij} would be the linear coefficients corresponding to the intensity of the intraction between atoms i and j per unit of contact area. Approximate values of the λ_{ij} coefficients (table I) can be derived adjusting the energies computed using equation (2) to empirical potentials such as those in the program package AMBER (Weiner et al., 1984) (Fig. 1). If also interactions of atom i with the solvent occur then the intensity of all non-covalent interacions of atom i (ΔG_i) will be from equations (1) and (2):

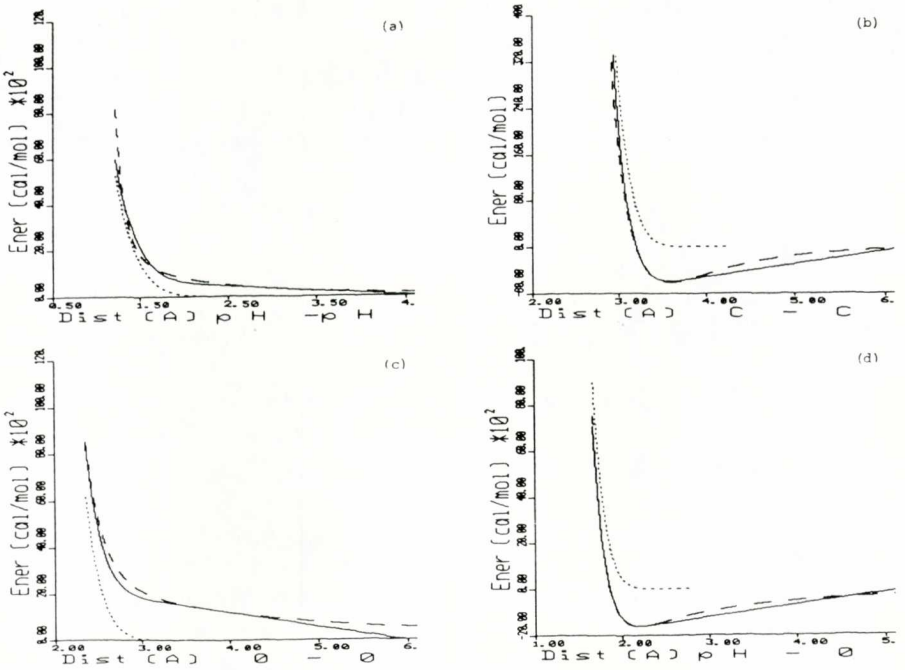

Figura 1. Non-bonded potentials for the Polar H-Polar H (a), $C(sp^3)$-$C(sp^3)$ (b), Polar O-Polar O (c) and Polar H-Polar O (d) non-covalent interactions. Potentials derived from the AMBER parameters (Weiner, Kollman, Case, Singh, Ghio, Alagona, Profeta & Weiner, 1984) with the Coulombic contribution included are indicated as a discontinous line. A dielectric constants of 4R as suggested by Whitlow & Teeter (Whitlow & Teeter, 1986) was used.

$$\Delta G_i = \Delta G_i^S + \Delta G_i^E = \Sigma_{j=0}\, \lambda_{ij}\, S_{ij}$$

where the summatory extends over all the interacting surfaces (S_{ij}) the atom i has. For j = 0, the corresponding interacting surface S_{i0} is the accessible surface of atom i and λ_{i0} coincides with the hydration coefficient g_i.

With the usual assumption that non-bonded interactions have additive contributions (Tanaka & Scheraga, 1976), the total non-bonded "empirical" energy of the protein molecule would be:

$$\Delta G = \Sigma_i\, \Delta G_i = \Sigma_i\, \Sigma_{j=0}\, \lambda_{ij}\, S_{ij} \qquad (3)$$

where the summation on i includes all protein parts whose interacions are considered. Again, for the sake of clarity, possible contributions due to the presence of short contacts in the structure have been omitted.

This unified description of the non-covalent interactions acting in a protein structure allows to evaluate the local relative stabilities of different protein conformations for a diversity of situations:

a) The stability of a given conformation is usually described (Dill, 1990) in terms of the denaturation free energy ($\Delta\Delta G^{\alpha}$) evaluated as the difference in free energy between the unfolded ($\Delta G_{unfolded}$) and folded (ΔG_{folded}) conformations under certain environmental conditions:

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

$\Delta\Delta G_i^{\alpha}$ would thus correspond to the stability of group i in that conformation.

Assuming that, in the unfolded state, the molecule is fully accessible to the solvent, which however, in most case, will be only partially of locally valid, we can write for group i:

$$\Delta\Delta G_i^{\alpha} \approx \Sigma_{j=1} \, (\lambda_{ij} - \lambda_{i0}) \, S_{ij} = \Sigma_{j=1} \, \alpha_{ij} \, S_{ij} \qquad (4)$$

As λ_{i0} and λ_{j0} are, in general unrelated (Eisenberg & McLachlan, 1986) α_{ij} is not symmetrical ($\alpha_{ij} \neq \alpha_{ji}$). The α_{ij} coefficients can be positive, meaning that the solvation is more favourable than the ij interacion, or negative in the opposite case.

b) To evalute how favourable is a particular interaction (λ_{ij}) for a given source surface i, that interaction could be compared with the most stable ineraction (λ_{iI}) the i surface can make. Then the global deviation from ideality ($\Delta\Delta G^{\beta}$) of all the non-covalent interactions in a protein structure can be expressed as:

$$\Delta\Delta G^{\beta} = \Delta G - \Delta G^{ideal} = \Sigma_i \, \Sigma_{j=0} \, (\lambda_{ij} - \lambda_{iI}) \, S_{ij} = \Sigma_i \, \Sigma_{j=0} \, \beta_{ij} \, S_{ij} \qquad (5)$$

where AG^{ideal} corresponds to an "ideal" conformation in which every surface i would be making only its most stable interacions (λ_{iI}). All the β_{ij} are, by construction, positives or zero when the observed interaction also corresponds to the best interaction of surface i. Again as λ_{ij} and λ_{ji} are, in general, unrelated (table 1), β_{ij} is not symmetrical ($\beta_{ij} \neq \beta_{ji}$).

c) The relative stability of two conformations (a and b) of a given protein can be expressed as:

$$\begin{aligned} \Delta\Delta G^{ba} &= \Delta\Delta G^{\alpha,b} - \Delta\Delta G^{\alpha,a} = (\Delta G^{b} - \Delta G_{unfolded}) - (\Delta G^{a} - \Delta G_{unfolded}) \\ &= \Delta G^{b} - \Delta G^{a} = \Sigma_i \, (\Sigma_{j=0} \, \lambda_{ij} \, (S_{ij}{}^{b} - S_{ij}{}^{a})) = \Sigma_i \, \Delta\Delta G_i^{ba} \end{aligned} \qquad (6)$$

Table 1. Derived lij coefficients[1] from the potentials[2].

	Non polar H	Polar H	C	O	N	S
Non polar 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

1. More types for different types of atoms could be used according to the analysis (de la Cruz, & Fita, in preparation)

2. The Amber potentials (Weiner, Kollman, Case, Singh, Ghio, Alagona, Profeta & Weiner, 1984) were used. The dielectric constant was taken as 4R as recomended by Whitlow & Teeter (Whitlow and teeter, 1986).

$\Delta\Delta G_i^{ba}$ would thus indicate the relative stability of the interactions of surface i in the two conformations a and b. As expected, the relative stability of the two conformations is independent of the unfolded state.

d) Finally, the relative stability between two given structures of different proteins a and b could be evaluated as:

$$\Delta\Delta G^{ba} = \Delta\Delta G^{\alpha,a} - \Delta\Delta G^{\alpha,a} = (\Delta G^b - \Delta G_{unfolded}) - (\Delta G^a - \Delta G_{unfolded})$$
$$= (\Delta G^b - \Delta G^a) + (\Delta G^a{}_{unfolded} - \Delta G^b{}_{unfolded}) \quad (7)$$

Thus the stabilities of both the denatured states and the given conformations are needed to compute the free energy balance between the two protein structures. Assuming again that the unfolded states are fully exposed for both protein we can evaluate equation (7) according to equation (4)

$$\Delta\Delta G^{ba} \approx \Sigma^b{}_{j=1}\, \alpha_{ij}\, S_{ij} - \Sigma^a{}_{j=1}\, \alpha_{ij}\, S_{ij} \quad (8)$$

This expression is specially suited when the two molecules are closely related, as would be the case for mutant proteins. In that case most terms appearing in equation (8) cancel. Minimization of equation (5) would allow to obtain the closest configuration to the ideal situation and thus the configuration with the lowest "empirical" (free) energy. As every term in equation (5) is positive or null the minimization of this expression can be done using a least squares algorithm. So we have added equation (5) to the well known Hendrickson & Konnert refinement package (Hendrickson & Konnert, 1980) as a new observational function ϕ_{int} for the non-covalent interactions that in the least squares formalism can be expressed as:

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

where m corresponds to the j interaction of surface i. Taking $\sigma^2(m) = 1/\beta_{ij}$ and $I_m = \sqrt{s_{ij}}$, we recover equation (5) since I_m^{ideal} would vanish for all possible interactions but the best in which case $\beta_{ij} = 0$ (de la Cruz & Fita, submitted).

Results

The structure of native Bovine Pancreatic Trypsin Inhibitor (BPTI) (Deisenhofer & Steigemann, 1975; Protein Data Bank reference: pdb4pti), with all the hydrogen atoms added in stereochemical ideal positions, was analysed according to the methodology presented in the previous paragraph (de la Cruz & Fita, submitted).

The relative stabilities of residues alkong the chain (Fig. 2) were tehn calculated using equation (4) for the atom types whose λ_{ij} coefficients are in table 1. These relative stabilities indicate that the structure of the native configuration appears to stabilize, with a few Kilocalories (per mol), the spatial disposition of almost every residue in the protein. The relative ideality is also shown (Fig. 2) with the corresponding atmic values calculated using equation (5) again for atom types in table 1. As expected the energies in the native configuration are well above those corresponding to an ideal situation however being, as indicated before, most of them stable.

A summary of the contat surfaces analysis in native BPTI (table 2a) and in an expanded BPTI (table 2b) are also presented. The expanded BPTI was obtained after 75 cycles of refinement using the algorithm implemented in the Hendrickson & Konnert package (de la Cruz & Fita, submitted). During the refinement every solvent atom interaction was considered softly attractive while the remaining non-bonded interactions were considered neutral, meaning that no explicit attraction nor repulsion was assigned to them. A clear increment of the solvent accessible surface in the expanded configuration (1813 Å) with respect to the initial native BPTI structure (862 Å) and also the related iminution of all the atom-atom interacting surfaces indicates that the expanded configuration corresponds to an "unfolded BPTI" with a large exposed surface. During this refinement the disulfide bridges of native BPTI were kept as covalent bonds and thus, however the important expansion obtained, the presence of these disulfide bridges are still limiting (strongly) the amount of accessible surface obtained for the extended BPTI structure obtained.

Expressions (3)-(5) provide a description of the non-covalent interactons acting in a protein structure in which only surfaces in close contact

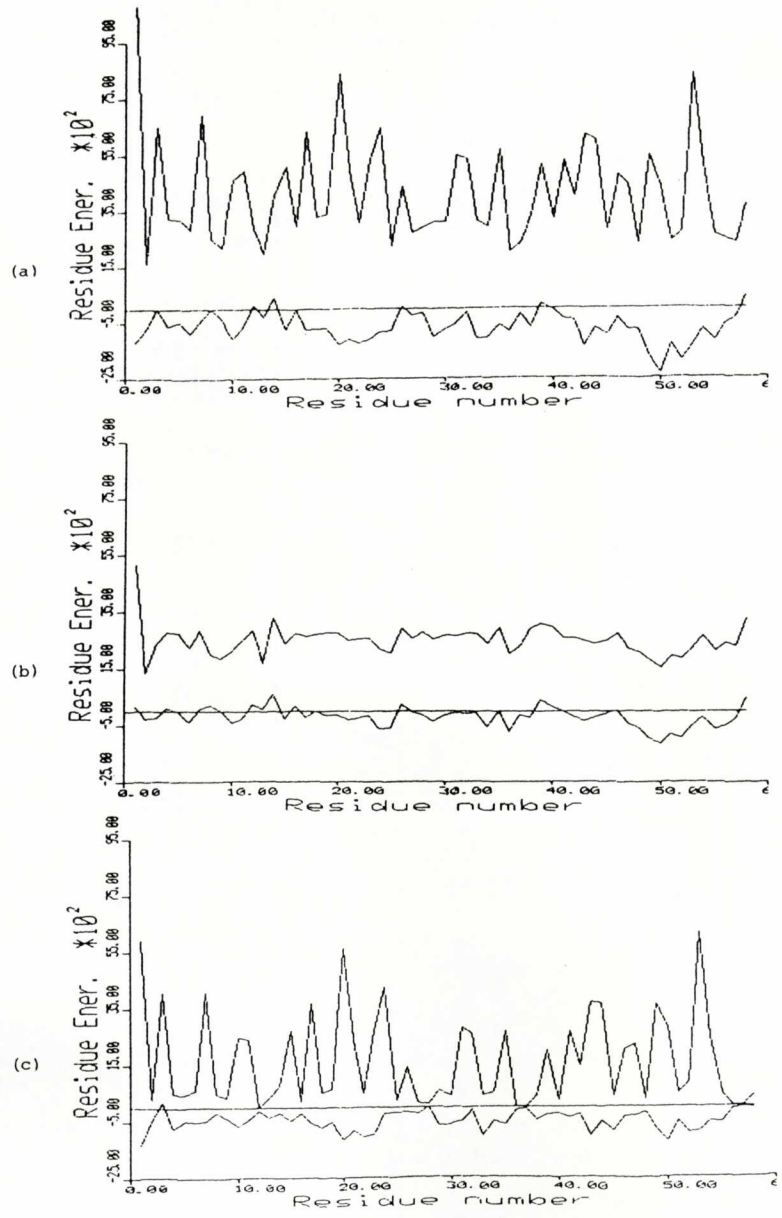

Figura 2. Stability energies of residues in BPTI. These values for each residue were obtained adding the values ofevery atom in the residue (a), of main chain atoms (b) and of side chain atoms (c). The atomic values were obtained using equation (4) in the test. The approximate λ_{ij} values were taken from (Table 1) and Eisenberg & McLachlan work (Eisenberg & McLachlan, 1986).

Table 2a. Summary[1] of contact surfaces[2] in native BPTI[3].

	Solvent	Apolar H	Polar H	C	N	O	S
Apolar H	415	609					
Polar H	210	368	48				
C	38	280	88	29			
N	33	78	14	29	1		
O	154	434	196	59	25	42	
S	12	92	13	14	5	18	6
Tot. Sur.[4]	862	2276	937	537	185	928	160

1. The number of atom types could be increased in a more detailed description (de la Cruz & Fita, in preparation).
2. The values of the surfaces listed correspond, for each kind of interaction between atoms of type i with atoms of type j, to $S_{ij} + S_{ji}$ if $j \neq 0$ (see the text).
3. The total surface of type i is obtained as $\Sigma_{j=1} S_{ij}$. The total accessible surface is taken as $\Sigma_i S_{i0}$; in both cases $i \neq 0$.
4. For covalent bond distances between non-hydrogen atoms the rms of the model was 0.023 Å.

Table 2b. Summary of contact surfaces in native BPTI[1].

	Solvent	Apolar H	Polar H	C	N	O	S
Apolar H	848	297					
Polar H	334	249	258				
C	159	166	40	34			
N	63	58	9	17	1		
O	347	226	102	54	9	24	
S	62	28	3	8	3	9	0
Tot. Sur.	1813	1872	762	478	160	771	113

1. The rms covalent bond distances bwtween non-hydrogen atoms in the model was 0.035 Å.

have significant contributions. This descripciton, being spatially localized, is well suited for a diversity of graphical analysis and displays that can highlight the peculiarities of the environment around each molecular group (de la Cruz, Reverter and Fita, submitted).

Acknowledgements

We would like to thank J. Ferrán, J. Reverter and J. Zaldivar for their stimulating suggestions and to Prof. J. A. Subirana for his constant sup-

port. This work was possible in part by a grant BT87 – 009 from the CICYT (Spain) and X. de la Cruz acknowledges a predoctoral fellowship from the Direcció General d'Universitats.

Abstract

The free energy of hydration and the potentials of mean force observed in protein sturctures can reproduced using accessible (Lee & Richards, 1971) and contact (Chothia, 1984) surfaces to describe solvent-atom and atom-atom non-covalent interactions respectively. A computer algorithm for calculating and optimizing the extent of these surface areas in a molecule that are accessible to the solvent or in contact with non-solvent atoms has been implemented as new restraints in a modified version of the least-squares refinement program of Hendrickson & Konnert (Hendrickson & Konnert, 1980). This information can thus be used either as added data during the crystallographic X-ray refinement or, independently, in the analysis and modelization of molecular structures. The algorithm appears very well suited for studying the structure and interactions of protein and nucleic acids.

REFERENCES

1. C. B. Anfinsen (1973) Science. *181*, 223.
2. C. Chothia (1974). Nature. *254*, 304.
3. J. Deisenhofer & W. Steigeman (1975) Acta Cryst. *B31*, 238-250.
4. K. A. Dill. (1990) Biochemistry. *29*, 7133-7155. D. Eisenberg & A. D. McLachlan (1986) Nature (London). *319*, 199-203.
5. J. L. Finney H. F. J. Savage. The Chemical Physics of Solvation. Part C; edited by R. R. Dogonadze, e. Kálmáan, A. A. Kornyshev & J. Ulstrup. pp. 603-663. Elsevier. The Netherlands.
6. I. Fita, A. Silva, M. Murthy & M. G. Rossmann (1986). Acta Cryst. *b42*, 497-515.
7. W. A. Hendrickson & J. H. Konnert (1980). Computing in Crystallography; edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 1301-1325. Bangalore; Indian Academy of Sciences.
8. J. Janin (1979). Nature. *277*, 491-492.
9. J. Janin, S. Miller & C. Chothia (1988). J. Mol. Biol., *204*, 155-164.
10. M. Karplus & J. A. McCammon (1983). Ann. Rev. Biochem. *53*, 263-300.
11. M. Karplus & G. A. Petsko (1990). Nature *347*, 631-639.
12. B. Lee & F. M. Richards (1971). J. Mol. Biol. *55*, 379-400.
13. G. J. Lesser & G. D. Rose (1990). Proteins: Structure, Function and Genetics. *8*, 6-13.
14. M. Levitt. (1982). Ann. Rev. Biophys. Bioeng. *11*, 251-271.
15. S. Miller, A. M. Lesk, J. Janin & C. Chothia. (1987) Nature. *328*, 834-836.
16. G. Némethy, M. S. Pottle & H. A. Scheraga (1983). J. Phys. Chem. *87*, 1883-1887.
17. T. Ooi, M. Oobatake, G. Némethy & H. A. Sscheraga (1987) P. N. A. S. *84*, 3086-3090.
18. F. M. Richards (1977) Ann. Rev. Biophys. Bioeng. *6*, 151-176.
19. T. J. Richmond & F. M. Richards (1978). J. M. Biol. *119*, 537-555.
20. G. D. Rose, R. H. Lee (1986) J. Biophys. *49*, 83-85.
21. R. S. Rowland, W. M. Carson & C. E. Bugg; in "Use of X-ray Crystallography in the Design of Antiviral Agents". Edited by W. G. Laver & G. M. Air. Academic Press, Inc. (1990), 261-281.
22. M. M. Santoro& D. W. Bolen. (1988) Biochemistry. *27*, 8063.

23. E. Silla, F. Villar, O. Nilsson, J. L. Pascual-Phnir & O. Tapia (1990). J. Mol. Graphisc. *8*, 168-172.
24. M. J. Sippl (1990) J. Mol. Biol. *213*, 859-883.
25. S. Tanaka & H. A. Scheraga (1976) Macromolecules. *9*, 945-950.
26. A. Warshel & S. Creighton (1989) Computer Simulation of biomolecular systems; edited by W. F. van Gunsteren & P. K. Weiner, pp. 120-138. ESCOM Science Publishers B. V., Leidenm, The Netherlands.
27. 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.
28. S. J. Weiner, P. A. Kollman, D. t. Nguyen, D. A. CASE (1986) J. of Computational Chemistry. *7*, 230-252.
29. M. Whitlow & M. M. Teeter (1986). J. Am. Chem. Soc. *108*, 7263-7172.