Universal geometrical factor of protein conformations as a consequence of energy minimization



2011 EPL 96 68005

(http://iopscience.iop.org/0295-5075/96/6/68005)

A Letters Journal Exploring the Frontiers of Physics

December 2011

EPL, **96** (2011) 68005
doi: 10.1209/0295-5075/96/68005

www.epljournal.org

# Universal geometrical factor of protein conformations as a consequence of energy minimization

Ming-Chya Wu[1,2,3(a)], Mai Suan Li[4], Wen-Jong Ma[2,5], Maksim Kouza[4,6] and Chin-Kun Hu[2(b)]

[1] *Research Center for Adaptive Data Analysis, National Central University - Chungli 32001, Taiwan*
[2] *Institute of Physics, Academia Sinica - Nankang, Taipei 11529, Taiwan*
[3] *Department of Physics, National Central University - Chungli 32001, Taiwan*
[4] *Institute of Physics, Polish Academy of Sciences - Al. Lotnikow 32/46, 02-668 Warsaw, Poland, EU*
[5] *Graduate Institute of Applied Physics, National Chengchi University - Taipei 11605, Taiwan*
[6] *Department of Physics, Michigan Technological University - Houghton, MI 49931, USA*

received 26 September 2011; accepted in final form 28 October 2011
published online 13 December 2011

PACS 87.14.E- – Proteins
PACS 87.15.A- – Theory, modeling, and computer simulation
PACS 87.15.-v – Biomolecules: structure and physical properties

**Abstract** – The biological activity and functional specificity of proteins depend on their native three-dimensional structures determined by inter- and intra-molecular interactions. In this paper, we investigate the geometrical factor of protein conformation as a consequence of energy minimization in protein folding. Folding simulations of 10 polypeptides with chain length ranging from 183 to 548 residues manifest that the dimensionless ratio $(V/A\langle r\rangle)$ of the van der Waals volume $V$ to the surface area $A$ and average atomic radius $\langle r\rangle$ of the folded structures, calculated with atomic radii setting used in SMMP (Eisenmenger F. *et al.*, *Comput. Phys. Commun.*, **138** (2001) 192), approach 0.49 quickly during the course of energy minimization. A large scale analysis of protein structures shows that the ratio for real and well-designed proteins is universal and equal to $0.491 \pm 0.005$. The fractional composition of hydrophobic and hydrophilic residues does not affect the ratio substantially. The ratio also holds for intrinsically disordered proteins, while it ceases to be universal for polypeptides with bad folding properties.



**Introduction.** – In recent decades, physical methods have been widely used to study properties and structures of biopolymers [1–3], including DNA [4,5], RNA [6], and protein [7–10]. Proteins assume specified conformations from their chemical compositions or sequences to develop biological activity and functional specificity. The corresponding three-dimensional (3D) structures are a consequence of inter- and intra-molecular interactions, in which energy minimization is the principle governing the folding tendency. In spite of various components involved in the interactions, there have been attempts to derive simple geometric factors from a variety of conformations, which can be either considered as a factor for structure validity or used as an effective constraint in folding simulation.

Geometric properties of protein molecules have been studied for more than three decades [11–13]. Among others, the Ramachandran plot [14] is a practical criterion widely used for improving the quality of NMR or crystallographic protein structures. In a polypeptide, the main chain N-$C_\alpha$ and $C_\alpha$-C bonds are relatively free to rotate, and can be respectively represented by two torsion angles. These angles can only appear in certain combinations due to steric hindrances, which define the allowed regions of the torsion angles for secondary structures in the plot.

Furthermore, it has been found that the mean volume of an amino acid in the interior of proteins *is very close to* that of the amino acid in crystals [11,12]. With the help of the Delaunary triangulation method, Liang and Dill [15] have reported that the protein packing is heterogeneous, and in terms of packing density, protein molecules may be either well-packed or loosely packed. Zhang *et al.* [16] showed that the packing density of single domain proteins decreases with chain length, which shares a generic feature of random polymers satisfying loose constraint in compactness.

Beside the Ramachandran plot and the packing density, which represent conclusions based on observations, there

(a) E-mail: mcwu@ncu.edu.tw
(b) E-mail: huck@phys.sinica.edu.tw

have been theoretical models introduced to simulate properties of protein geometric structures. For example, Banavar and Maritan [17] have introduced the effective backbone tube model to analyze the secondary structures of proteins under the constraint of minimum energy and showed that the tube has an effective radius of 2.7 Å.

When a polypeptide folds, the hydrophobic effects cause non-polar side chains to cluster together in the protein interior or interface, whereas polar side chains tend to maximize the contacts with outer solvent molecules. The stability of the system is partially due to the burial of the non-polar residues, and can be measured by the loss of the solvent accessible surface area [18–21]. An atom or group of atoms is defined as accessible if a solvent molecule of specified size, generally water, can be brought into van der Waals contact. The solvent accessible surface is then simply defined as the surface traced out by the center of a probe sphere, which represents the solvent molecule, as it rolls over the van der Waals surface of the protein [20,22]. Hence, volume and surface area are suitable parameters to characterize the geometrical conformation of protein.

**Methods.** –

*Folding simulation.* We used the SMMP package [23,24] for protein folding simulation and simulated annealing, as well as canonical Monte Carlo method, to generate folded structures. Starting with a polypeptide in a solvent, the SMMP searches the lowest energy conformation by utilizing the energy function

$$E_{tot} = E_{LJ} + E_{el} + E_{hb} + E_{tors}, \tag{1}$$

where

$$E_{LJ} = \sum_{j>i} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right), \tag{2}$$

$$E_{el} = 332 \sum_{j>i} \frac{q_i q_j}{\varepsilon r_{ij}}, \tag{3}$$

$$E_{hb} = \sum_{j>i} \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right), \tag{4}$$

$$E_{tors} = \sum_{n} U_n \left[ 1 \pm \cos(k_n \phi_n) \right]. \tag{5}$$

Here $r_{ij}$ is the distance in Å between atoms $i$ and $j$. $A_{ij}$, $B_{ij}$, $C_{ij}$, and $D_{ij}$ are parameters of the empirical potentials. $q_i$ and $q_j$ are the partial charges on the atoms $i$ and $j$, respectively, $\varepsilon = 2$ is the dielectric constant of the protein interior space. The factor 332 in eq. (3) is used to express the energy in kcal/mol. $U_n$ is the energetic torsion barrier of rotation about the bond $n$ and $k_n$ is the multiplicity of the torsion angle $\phi_n$ [17]. The input file for SMMP is a sequence of amino acids and the output file is in the Protein Data Bank (PDB) format [25]. The protein-solvent interactions were implemented with the implicit water solvation by selecting type 1 solvent in the SMMP *main.f* program. All parameters needed for the simulation have been self-contained in the SMMP package.

*Calculation of volume and surface area.* To compute the volume $V$ and surface area $A$ of the polypeptide in the course of folding simulation, we used the ARVO package [26] developed based on analytic equations [22]. ARVO can calculate $V$ and $A$ of a system of $N$ atoms, which can overlap in any way. The main idea of the algorithms of ARVO is converting computation of volume and surface area of overlapping spheres as surface integrals of the second kind over closed regions. Using stereographic projection, one can transform the surface integrals to a sum of double integrals which are then reduced to curve integrals [22,26]. It has been shown that the van der Waals surface areas [27] computed by the GETAREA module in FANTOM package [28,29] and ARVO module are consistent [26]. Comparing with programs implementing different algorithms and approximations to describe geometrical properties of atomic groups, the differences among the computed surface area by VOLBL [30], GEPOL [31,32] and ARVO [26] are less than 1%, and the differences among the computed volumes are about 2% (see refs. [26] and [33] for detailed discussions). On the basis of analytical method, the accuracy of the computation of volume and surface area of protein molecules by using ARVO is superior to numerical integration which always contains numerical errors [33].

The input file for ARVO contains the coordinates $(x_i, y_i, z_i)$ of the center and radius $r_i$ of all $N$ atoms in the system, where $1 \leqslant i \leqslant N$. The atoms can overlap in any way. To calculate van der Waals surface area $A$ and volume $V$ of a PDB protein structure, we used the coordinates of carbon (C), nitrogen (N), oxygen (O), and sulfur (S) of the PDB data and van der Waals radii of C, N, O, and S as input data. According to the conventional parameter settings in protein folding simulations [23,24,34–37], N atom has (van der Waals) radius 1.55 Å, C atom has radius 1.55 Å, S atom has radius 2.00 Å, and O atom has radius 1.40 Å. The relatively smaller radius of hydrogen (H) atom is neglected; a water (solvent) molecule is represented by an O atom with radius 1.40 Å. The radii of these atoms at the atomic level are determined by the densities of the electron cloud, and they are self-consistent with other physical quantities used in the SMMP simulation [23,24]. Further, to calculate solvent accessible surface area $A_s$ and related volume $V_s$ of the protein structure, we added the radius of the solvent 1.40 Å to van der Waals radii of C, N, O, and S, *i.e.* the effective radii of C, N, O, and S are 2.95 Å, 2.95 Å, 2.80 Å, and 3.40 Å [22,26], respectively. The average atomic radius $\langle r \rangle$ and average effective radius $\langle r_s \rangle$ of folded structures are calculated using these radii.

**Results and discussions.** –

*$V/A\langle r \rangle$ ratio for the van der Waals volume $V$ and surface area $A$ and average atomic radius $\langle r \rangle$.* The ratio $R = V/A\langle r \rangle$ and the total energy are computed in the time course of simulation. In all cases of our simulation, the final energy using canonical Monte Carlo method is lower than using simulated annealing. The results of 10 small proteins

Table 1: The $V/A\langle r\rangle$ ratio for 10 typical proteins structures. $R'$ is the $V/A\langle r\rangle$ ratio of the structure with a randomly chosen configuration by the SMMP package [23,24], $R''$ is for the final structure after performing the folding simulation, the subscript "$a$" stands for simulated annealing and "$c$" for canonical Monte Carlo, and $R$ is for the structure from PDB.

| PDB | $N$ | $R'_{(a)}$ | $R''_{(a)}$ | $R''_{(c)}$ | $R$ |
|---|---|---|---|---|---|
| 1HP9 | 183 | 0.5694 | 0.4912 | 0.4875 | 0.5047 |
| 1KDL | 193 | 0.5364 | 0.4867 | 0.4863 | 0.4998 |
| 1GCN | 246 | 0.5480 | 0.4917 | 0.4875 | 0.4946 |
| 1VII | 295 | 0.5560 | 0.4876 | 0.4878 | 0.5058 |
| 2PLH | 330 | 0.6280 | 0.4864 | 0.4857 | 0.4954 |
| 2OVO | 418 | 0.5549 | 0.4981 | 0.4891 | 0.4933 |
| 1PGB | 436 | 0.5385 | 0.4866 | 0.4878 | 0.4891 |
| 1HPT | 440 | 0.5726 | 0.4879 | 0.4871 | 0.4928 |
| 1UOY | 452 | 0.5414 | 0.4896 | 0.4872 | 0.4945 |
| 1UTG | 548 | 0.5785 | 0.4780 | 0.4858 | 0.4896 |

(with $183 \leqslant N \leqslant 548$) (table 1) reveal that $R$ approaches to $\approx$0.49 as the energy decreases, while the resultant structures are not necessary close to native structures. It turns out that the energy minimization criterion is likely connected with the geometric conformation defined by the ratio $R \simeq 0.49$.

To confirm that the ratio $R \simeq 0.49$ is relevant, we have tested 743 PDB structure data from the Protein Culling Server [38], in which only X-ray data with high resolution have been selected. The ratio is found to be $R = 0.491 \pm 0.005$. To determine a reasonable tolerance for the ratio, we have also tested a larger database from the Protein Data Bank. Totally 31059 PDB entries deposited at the Protein Data Bank in June 2005 have been downloaded for the test. After excluding non-proteins, such as DNA and RNA, and problematic structures in which only $\alpha$ carbons are included, there are finally 28664 protein structures involved in statistics. In our analysis, both X-ray and NMR data have been used. For NMR data consisting of more than one model, we selected the first model which is considered as the most accurate one or is an average of the models. We plotted the dependence of van der Waals surface area $A$ and volume $V$ on the total number of (C, N, O, and S) atoms $N$ in fig. 1(a) which shows that $A$ and $V$ increase linearly with $N$. The linear correlation between the volume $V$ and the number of atoms $N$ or area $A$ has been found by Lorenz *et al.* [39] by using the Monte Carlo studies with the model of clusters of random uncorrelated spheres. Similar results of the linear relations have also been discussed by Liang and and Dill [15] with 636 protein structures. The result in fig. 1(a) provides a more solid demonstration from the basis of a larger database. Furthermore, we plotted the distribution of $R = V/A\langle r\rangle$ as histograms in blue (solid line) in fig. 1(b), which locates in a very narrow interval centered at $R = 0.4910$.

$R \approx 0.491$ implies that one cannot imagine a protein as a chain of small spheres because in this case we would have $R_c = (4\pi\langle r\rangle^3/3)/4\pi\langle r\rangle^2\langle r\rangle = 1/3 \approx 0.333$. However,

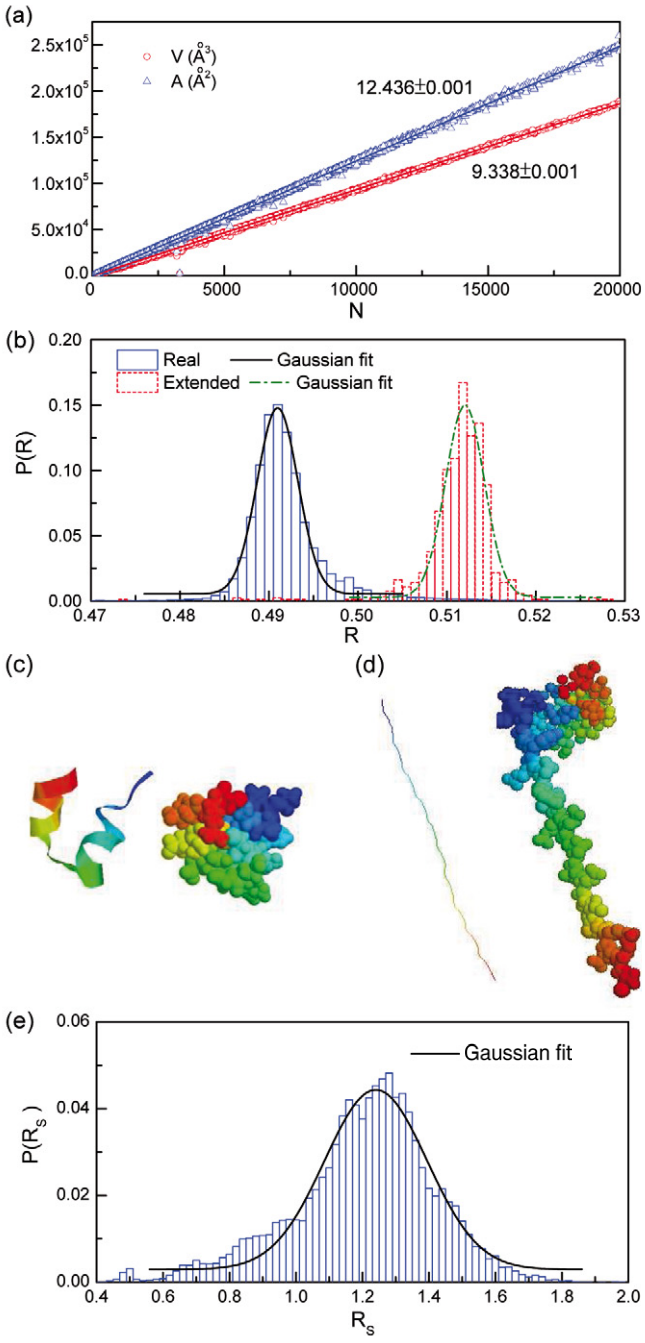


Fig. 1: (Colour on-line) (a) Dependence of $V$ and $A$ on $N$ for 28664 PDB protein structures. The numbers indicate the slopes. (b) The distribution $P(R)$ (blue, solid line) for the structures shown in (a). The Gaussian fit with the maximum located at $R \simeq 0.491$, with $y = y_0 + \frac{S}{w\sqrt{\frac{\pi}{2}}} \exp\left[-\frac{2(x-x_c)^2}{w^2}\right]$, $y_0 = 0.0053, x_c = 0.4910, w = 0.0046$, and $S = 0.0008$. The estimation of the fitting is adjusted $\Re^2 = 0.984$ for 932 artificial extended structures is shown in red (dotted line). The green (dash-dotted) line refers to the Gaussian fit with maximum at $R \approx 0.5120$. (c) A typical compact PDB structure (PDB code: 1VII). (d) An extended structure obtained by Swiss-pdb viewer 3.7 (SP5). (e) The distribution of $R_s$ for 28236 proteins using the probe sphere with radius of 1.4 Å. The maximum of the Gaussian fit is located at $R_s \approx 1.2402$.

the result $R \approx 0.491$ might be understood qualitatively by considering that a protein consists of tubes of radius $\langle r \rangle$. There is a tube to represent the backbone of the protein; there are also some tubes to represent side chains of the protein. The total length of tubes is $l \sim N$. Using $V \approx \pi \langle r \rangle^2 l$ and $A \approx 2\pi \langle r \rangle l$ we obtained $V/A\langle r \rangle = 1/2 \approx 0.5$ which is consistent with our numerical result. The linear dependence of $V$ and $A$ on $l \sim N$ is supported by fig. 1(a). It is worth noting that the linear correlation is independent of the settings of the radii of atoms. If other radii are used, the linear relation remains but the ratio is different. The ratio derived from the average of an ensemble of 715 PDB protein chains (selected by Protein Sequence Culling Server [38]), using Richard's parameters [18], is $V/A\langle r \rangle = 0.5589 \pm 0.0114$. Similarly, using the Protori radii [40], the result is $V/A\langle r \rangle = 0.5288 \pm 0.0113$. The relation $V/A\langle r \rangle \approx 1/2$ approximately holds in the two cases.

To clarify the relation between the ratio $R \simeq 0.491$ and the compactness of native structures, we computed $R$ for artificial extended structures of protein molecules. The extended structures are obtained by setting all torsion angles of existing 3D structures from PDB equal to 180°, using the Swiss-pdb viewer 3.7 (SP5) (`http://www.expasy.org/spdbv/`). A typical compact PDB structure and extended structure are shown in fig. 1(c) and fig. 1(d), respectively; the latter is similar to those obtained from mechanical unfolding of proteins studied in refs. [41,42]. The histograms in red (dotted line) in fig. 1(b) show the distribution of $R$ for 932 artificial structures. Its maximum locates at $R \approx 0.5120$ which is higher compared to real protein structures. This interesting result confirms that the value $R \approx 0.491$ comes from the requirement for the formation of compact native conformations as a result of energy minimization.

Further, direct comparisons of volumes and surface areas for real and extended structures show that from an extended structure to a real structure, there is a small change (increase or reduction) in volume while there is usually a large increase in surface such that the ratio changes from 0.512 to 0.491. All of these indicate that the larger ratio of extended structure is attributed to non-physical geometrical properties, such as loosely connections of monomers, and unbalance of electrostatic interactions among monomers, and interactions between monomers and water molecules. It should be noted that both real and artificial structures satisfy the requirements imposed by the Ramachandran plot, but only the former has protein-like properties. Thus, $R \approx 0.491$ can serve as a useful factor for selecting three-dimensional protein-like structures. In addition, we have also found that the beta structures extracted from PDB protein structures have smaller $R = 0.4808$ in comparison with the helixes ($R = 0.4859$), as shown in fig. 2(a). Whereas the ratios for individual secondary structures are different, a protein molecule as a whole is a self-organized geometric unit, which blends various secondary structures to form a properly folded 3D structure. In contrast to secondary structures, tertiary and quaternary structures then have the universal property of $R \simeq 0.491$ regardless of their details. Defining the relative beta/helix content of a protein as a number of amino acids belonging to beta strands/helix structures divided by its total number of residues, we found that, as shown in figs. 2(b) and (c), there is no correlation between $R$ and beta- as well as helix-content as the correlation level for the linear fits is very low (0.005) for both cases.

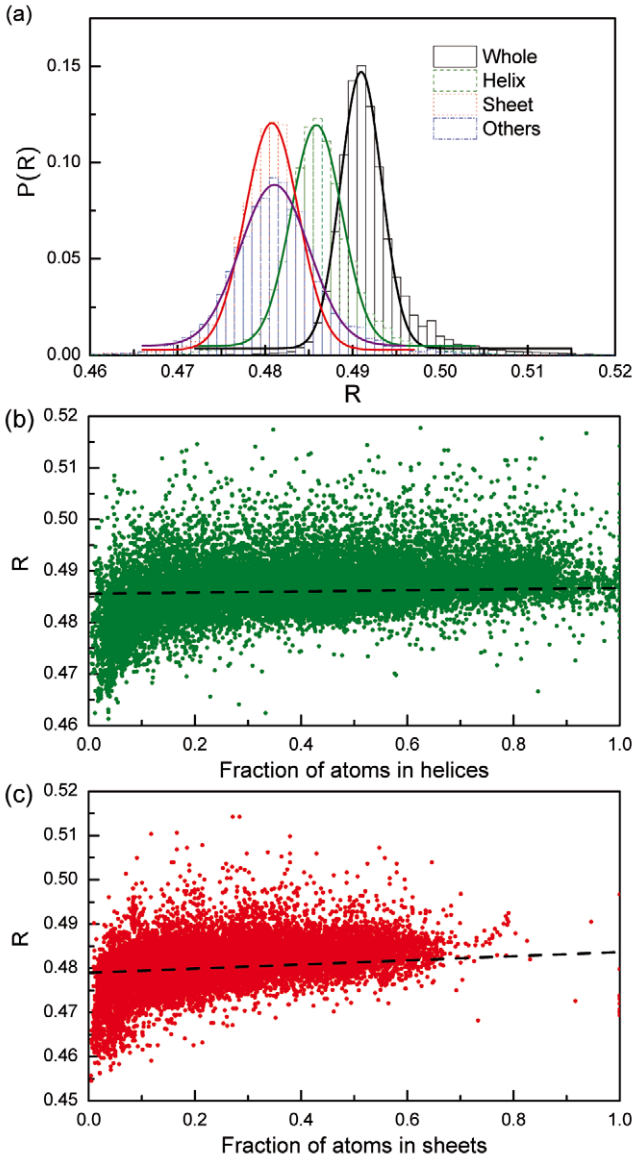


Fig. 2: (Colour on-line) (a) Probability density function of $R$ statistics for whole protein molecules (28664 samples, Gaussian distribution centered at $R = 0.4910$), helix structures (extracted from 26040 samples, $R = 0.4859$), sheet structures (24537 samples, $R = 0.4808$) and other structures (25513 samples, $R = 0.4811$). (b) $R$ as a function of the fraction of atoms in helix structures. The slope of the linear fit (black dashed line) is 0.0001, and the correlation level is 0.005. (c) $R$ as a function of the fraction of atoms in sheet structures. The slope of the linear fit (black dashed line) is 0.0002, and the correlation level is 0.005.

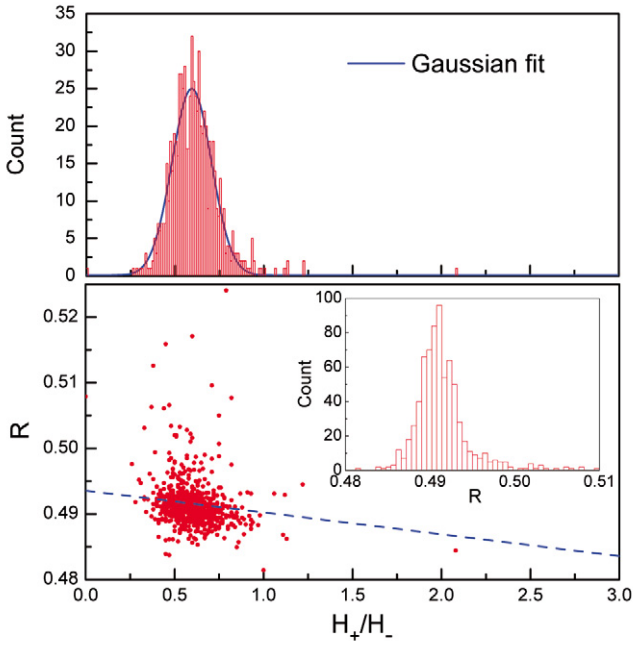


Fig. 3: (Colour on-line) The upper panel shows the distribution of the ratio of hydrophilic ($H_+$) and hydrophobic ($H_-$) amino acids in a molecule for 723 protein structures (from Protein Sequence Culling Server [38]), based on the Kyte-Doolittle scale [43]. The Gaussian fit is centered at 0.594. The lower panel shows $R$ as a function of $H_+/H_-$. The slope of the linear fit (blue dashed line) is $-0.0033$, and the correlation level is 0.2. The inset shows the distribution of $R$ for the 723 protein structures.

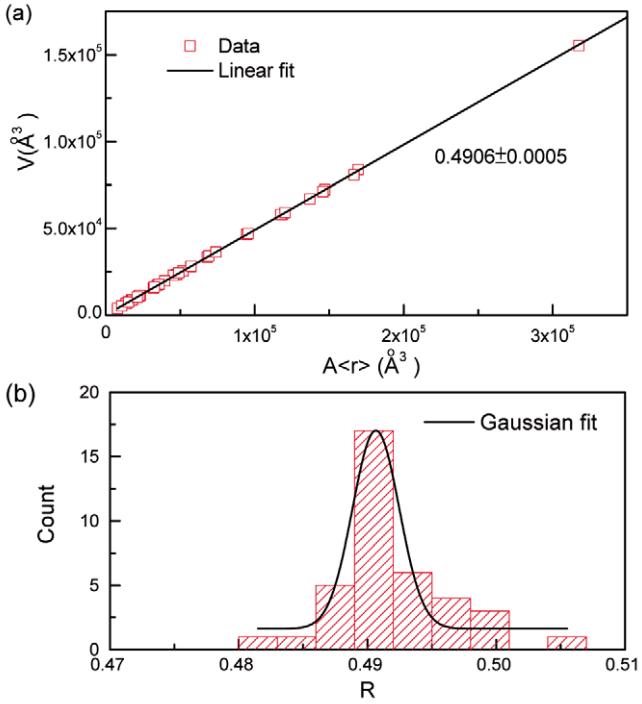


Fig. 4: (Colour on-line) (a) Volume $V$ as a function of surface area $A$ for 38 disordered proteins [44]. (b) The histogram of the structures shown in (a). The Gaussian fit has a maximum located at $R \simeq 0.4906$.

*$V_s/A_s\langle r_s\rangle$ ratio for solvent accessible volume $V_s$ and surface area $A_s$ and average effective radius $\langle r_s\rangle$.* In order to compare the distributions of $R$ (with zero radius of solvent) and $R_s \equiv V_s/A_s\langle r_s\rangle$ (with radius of solvent 1.4 Å), we calculated the distributions of $R$ and $R_s$ for 28236 protein structures from PDB. The histogram of $R$ is not shown because it is similar to the larger set of 28664 protein structures (fig. 1(b)). The distribution of $R_s$ (fig. 1(e)) has a maximum at $R_s \approx 1.2402$.

*$V/A\langle r\rangle$ ratio and hydrophobicity of amino acids.* Consider a polypeptide chain consisting of a sequence of amino acids with different hydrophobicities. The hydrophobic condensation drives the polypeptide chain toward a conformation with lower free energy. This is achieved by burying hydrophobic contents into the interior and residing polar monomers on the surface in contact with water. This process involves not only the regulation of the connections between monomers but also compensations of volume and surface area. According to the statistics shown in fig. 3 for 723 protein structures (from Protein Sequence Culling Server [38]), the ratio of hydrophilic and hydrophobic amino acids ($H_+/H_-$) of proteins in the Kyte-Doolittle scale [43] is generally in a narrow range with respect to the variable range of $H_+/H_-$, suggesting that the universality of $R$ is probably a consequence of compositions of hydrophilic and hydrophobic amino acids in a protein. The linear fit for $R$ as a function of $H_+/H_-$ (lower part of fig. 3) gives the correlation level of 0.2. Since this level is notably lower than 0.5, there is no correlation between these two quantities. This is not unexpected because $R$ varies in a very narrow interval. For this reason, one can show that $R$ does not correlate with individual values of $H_+$ and $H_-$. Furthermore, folding simulations of ten polypeptides with fixed $H_+/H_-$ and randomized sequences show that the averages of the ratios are $\langle R'\rangle = 0.548 \pm 0.013$ for initial structure and $\langle R''\rangle = 0.490 \pm 0.001$ for final structure after energy minimization. This implies that the ratio is not only the property of disordered protein, but is also that of random copolymers.

*$V/A\langle r\rangle$ ratio for intrinsically disordered proteins.* It is also of interest to study the ratio for the intrinsically disordered proteins which usually lead to misfolding [44,45]. We have calculated $R$ for 38 protein structures [44] and found that the ratio is $0.4906 \pm 0.005$ (see fig. 4), which is within the tolerance determined by the ensemble of 28664 PDB protein structures. Thus, the ratio holds once a polypeptide folds to a compact structure no matter of its species.

**Summary and conclusions.** – In summary, we have found a universal ratio of the van der Waals volume to the surface area and average atomic radius of folded structures $R = 0.491 \pm 0.005$ for native protein structures, including intrinsically disordered proteins. We have studied

the connection between the energy minimization and the geometric conformation by monitoring the ratio $R$ during folding simulations using the SMMP package [23,24]. Our results reveal that $R \simeq 0.491$ should be somewhat related to the energy global minimum of protein molecules. This result can be imposed as a rule in searching for native conformations in folding simulations using protein sequences. $R \approx 0.491$ can also serve as a necessary condition for checking the validity of PDB data and designing protein-like sequences.

It is well known that hydrophobic residues are buried in the core of proteins and the van der Waals volume should be, therefore, proportional to the number of such residues. The van der Waals area should linearly depend on the number of hydrophilic residues, which have tendency to reside on the protein surface. Thus, the universality of $R$ is probably a consequence of the fact that the ratio of the hydrophobic and hydrophilic amino acids of proteins is roughly a constant.

Here we should emphasize that $R \simeq 0.491$ does not correspond to a unique conformation, but it confines molecular conformations in a folding simulation from vast possibilities to a smaller space. It excludes improperly folded structures which are characterizable by such geometrical properties and is beneficial for the reduction of simulation time. One possible implementation of this property shall be in the calculation of surface energy associated with the solvent access area. A preliminary test of the $V/A\langle r\rangle$, working as a filter, can be performed before next update step in simulations. Other independent factors can work together to define the conformation to have a native-like structure.

* * *

This work was supported by Taiwan Grants NSC 96-2112-M-008-021-MY3, 96-2911-M-001-003-MY3, 97-2627-B-008-004, 98-2627-B-008-004, 99-2627-B-008-002, 100-2112-M-008-003-MY3, Poland-Taiwan NSC Grant 100-2911-I-001-507, NCTS (North), AS-95-TP-A07, and the Polish Grant No. 202-204-234.

REFERENCES

[1] Glaser R., *Biophysics* (Springer-Verlag, Berlin) 2001.
[2] Huang K., *Lectures on Statistical Physics and Protein Folding* (World Scientific, Singapore) 2005.
[3] Waigh T. A., *Applied Biophysics* (John Wiley & Sons, England) 2007.
[4] Peng C.-K. *et al.*, *Nature*, **356** (1992) 168.
[5] Lando D. Y. *et al.*, *EPL*, **91** (2010) 38003.
[6] Mamasakhlisov Y. S. *et al.*, *Phys. Rev. E*, **75** (2007) 061907.
[7] Hyman B. T. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **92** (1995) 3586.
[8] Okamoto Y., *Physica A*, **254** (1998) 7.
[9] Dokholyan N. *et al.*, *Fold. Des.*, **3** (1998) 577.
[10] Gevorkian S. G. *et al.*, *EPL*, **95** (2011) 23001.
[11] Chothia C., *Nature*, **254** (1975) 304.
[12] Chothia C., *Annu. Rev. Biochem.*, **53** (1984) 537.
[13] Richards F. M., *J. Mol. Biol.*, **82** (1974) 1.
[14] Ramachandran G. N. *et al.*, *J. Mol. Biol.*, **7** (1963) 95.
[15] Liang J. and Dill K. A., *Biophys. J.*, **81** (2001) 751.
[16] Zhang J. *et al.*, *J. Chem. Phys.*, **118** (2003) 6102.
[17] Banavar J. R. and Maritan A., *Rev. Mod. Phys.*, **75** (2003) 23.
[18] Lee B. and Richards F. M., *J. Mol. Biol.*, **55** (1971) 379.
[19] Richards F. M., *Annu. Rev. Biophys. Bioeng.*, **6** (1977) 151.
[20] Connolly M. L., *Science*, **221** (1983) 709.
[21] Stephanie M. E. T. *et al.*, *J. Am. Soc. Mass Spectrom.*, **17** (2006) 1490.
[22] Hayryan S. *et al.*, *J. Comput. Chem.*, **26** (2005) 334.
[23] Eisenmenger F. *et al.*, *Comput. Phys. Commun.*, **138** (2001) 192.
[24] Eisenmenger F. *et al.*, *Comput. Phys. Commun.*, **174** (2006) 422.
[25] Berman H. M. *et al.*, *Nucl. Acids Res.*, **28** (2000) 235.
[26] Buša J. *et al.*, *Comput. Phys. Commun.*, **165** (2005) 59.
[27] Labanowski J. K., *Molecular Modeling* (1996), posted at `http://www.ccl.net/cca/documents/molecular-modeling/node1.html.shtml`.
[28] Fraczkiewicz R. and Braun W., *J. Comput. Chem.*, **19** (1998) 319.
[29] Freygerg B. V. and Braun W., *J. Comput. Chem.*, **14** (1993) 510.
[30] Liang J. *et al.*, *Proteins: Struct. Funct. Genet.*, **33** (1998) 1.
[31] Silla E. *et al.*, *J. Mol. Graphics*, **8** (1990) 168.
[32] Silla E., Tun I. and Pascual-Ahuir J. L., *J. Comput. Chem.*, **12** (1991) 1077.
[33] Buša J. *et al.*, *J. Comput. Chem.*, **30** (2009) 346.
[34] Ghulghazaryan R. G. *et al.*, *J. Comput. Chem.*, **28** (2007) 715.
[35] Hansmann U. H. E. and Okamoto Y., *J. Comput. Chem.*, **14** (1993) 1333.
[36] Hayryan S. *et al.*, *J. Comput. Chem.*, **22** (2001) 1287.
[37] Lin C.-Y. *et al.*, *Proteins: Struct. Funct. Genet.*, **52** (2003) 436.
[38] Wang G. and Dunbrack R. L. jr, *Bioinformatics*, **19** (2003) 1589 . The website of the protein sequence culling server is `http://dunbrack.fccc.edu/pisces/`. We used version cullpdb_pc20_res1.6_R0.25_d040724_chains743.
[39] Lorenz B. *et al.*, *J. Phys. A: Math. Gen.*, **26** (1993) 4711.
[40] Tsai J. *et al.*, *J. Mol. Biol.*, **290** (1999) 253.
[41] Li M. S. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 93.
[42] Li M. S. *et al.*, *Biophys. J.*, **92** (2007) 547.
[43] Kyte J. and Doolittle R. F., *J. Mol. Biol.*, **157** (1982) 105.
[44] Mészáros B. *et al.*, *J. Mol. Biol.*, **372** (2007) 549 . There are 39 PDB structures in table 1, while the entry "1F83" has been obsoleted by the Protein Data Bank.
[45] Mittag T. and Forman-Kay J. D., *Curr. Opin. Struct. Biol.*, **17** (2007) 3.