THE LONG-RANGE PROPERTIES OF PROTEIN STRUCTURE AND MOBILITY

V.V.MORARIU, OANA ZAINEA

National Institute for R&D of Isotopic and Molecular Technologies,
71–103, Donath street, Cluj-Napoca, 400293, Romania, e-mail: *vvm@L40.itim-cj.ro*

Abstract. A correlation investigation was carried out on the series atomic coordinates (*MPV*) and temperature factor (T_f) respectively for protein main chains. The *MPV* series revealed resemblances to the corresponding T_f series in the case of human hemoglobin and HIV-1 protease. Each minor or major peak in a series had a corresponding peak in the related series. This brings a qualitative evidence for the connection of the two parameters. The series were further subject to spectral and detrended fluctuation analysis. They were characterized in terms of scaling exponents, as the series showed long-range correlation properties. The *MPV* series are stronger correlated structures and much less sensitive to ligand binding than the T_f series. The T_f series are comparatively much more sensitive to the same factors. Thus the long-range correlation properties of the protein main chain structure and mobility reveal two complementary properties of protein molecules. While structure imparts constancy to the system, the atomic mobility represents function yet they are intimately connected.

Key words: Protein, atomic coordinates, temperature factors, long-range correlation.

INTRODUCTION

The Protein Data Bank (PDB) contains the atomic coordinates and the temperature factor (or atomic mobility) of a large number of proteins whose structures were determined by either X ray crystallography or NMR techniques. The purpose of this project was to establish the correlation properties of both the atomic structure and the atomic mobilities of the protein main chains in order to answer the following questions: a) What is the relationship between the atomic structure and mobility? and b) What is the sensitivity of the atomic structure and mobility to ligand binding? The goal of the study belongs to the basic question regarding the relationship between structure, mobility and function of the proteins. This is at present regarded as an unsolved problem.

Received July 2005;
In final form October 2005

ROMANIAN J. BIOPHYS., Vol. 15, No. 1-4 P. 3–8, BUCHAREST, 2005

We regarded the atomic structure (atomic coordinates) and mobility (temperature factor, T_f) data of a protein backbone as a series where the *i*-th atom position number replaced time. Then $T_f = f(i)$ where i =1…n, for n atoms in the protein backbone. The order is kept according to the natural succession of the atoms in the backbone. Using this procedure it was found that T_f series obey a power law, $P = 1/f^{\beta}$, where *P* represents the spectral power, *f* is the frequency and β is the scaling exponent characterizing the long-range correlation (1–2).

In this study we applied a similar procedure on the structural data of the main chain atoms and compared to the correlation characteristics of T_f data. We selected, for this report, a HIV-1 protease which is a protein belonging to the beta-class of proteins. Calculations were performed on other types of proteins as well. Our protein case was investigated in the presence of various ligands so that the effect of these factors on the correlation properties could be studied at the same time.

METHODS

The x, y, z, atomic coordinates of the protein main chains were extracted from the Protein Data Bank (PDB) for the selected protein. We choose the magnitude of the position vector (*MPV*) as a characteristic parameter for the position of an atom, which was suitable to handle in our correlation analysis. The most direct approach is to calculate the *MPV* with respect to the origin of the crystallographic unit cell coordinate system. We should bear in mind that a protein structure is referred to a unit cell which is characterized by the crystal cell lengths *a*, *b*, and *c* and angles α, β and γ. These data are available for each protein in PDB. Therefore the general definition of *MPV* in a non-rectangular Cartesian coordinates system is $MPV = \sqrt{x^2 + y^2 + z^2 + 2xy\cos\gamma + 2xz\cos\beta + 2yz\cos\alpha}$. The x, y and z coordinates have an origin which is placed in a side corner of the unit cell. Further the series of temperature factors for the same main chain atoms were extracted from PDB. Both kind of series were subjected to Fast Fourier Transform and the slope of the double logarthmic plot was the scaling exponent β. This exponent reflect both the stationary and nonstationary contributions to the overall long-range correlation.

Further we have used the Detrended Fluctuation Analysis (DFA) of the same series to measure the scaling exponent α which is a measure of the long-range correlation due to the stationary contribution of the series (3). The relationship in a stationary series between the two scaling exponents is: $\beta = 2\alpha - 1$. Obviously, when a nonstationary series of data is considered then, $\beta_{nonstat} > \beta_{stat}$ and the $\beta = 2\alpha - 1$ relatioship is no longer valid. This is why we propose a new way to characterize a nonstationay series by using both the α and β scaling exponents. So each case is characterized by the pair of α and β values. At the same time this may be regarded as a simple diagnostic tool for a series: If $\beta = 2\alpha - 1$ relationship is verified, then

the series is stationary. While in the case of many time series the nonstationary character is regarded as an unwanted contribution, the nonstationary character of the structure and mobility series of proteins represent an intrisic property and therefore deserves full attention.

RESULTS AND DISCUSSION

An example of a power spectrum for the atomic coordinates of the main chain atoms of a protein is presented in Fig.1. Similar spectra can be obtained for any series of data of atomic coordinates or temperature factors. As the spectrum is linear in a double log plot, the long-range correlation can be infered for all these cases. The interesting point is that slope of the spectra, i.e. the α and β scaling exponents behave in a different way for the same protein. The following distinctions were found: a) As the experimental value of β is greater than the calculated value of a hypothetical β (β = 2α – 1, where α is the result of the DFA corresponding to a stationary series) it follows that the series are nonstationary; b) The α and β exponents of the protein structure is insensitive to ligand binding and, c) The same exponents are very sensitive to ligand binding for the temeprature factor series, i.e. the atomic mobility.

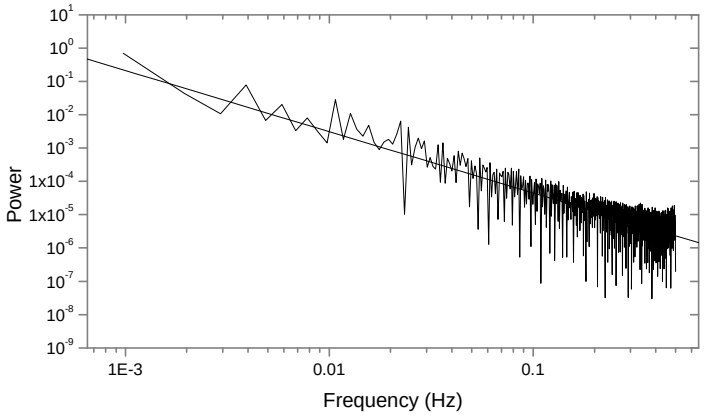

Fig. 1. Power spectrum of the Magntitude of Position Vector for the main chain atoms of HIV-1 protease dimer (PDB cod:1qbs). The linear fitting has a slope –1.83702±0.06613.

This is illustrated in Fig. 2 for the same protein in the presence of various ligands. We can easily notice that the position of each case on a β = *f*(α) plot is very similar for the *MPV* data while the data are spread along a straight line for the

atomic mobility. In other words the long-range correlation exponents of the structure is little sensitive to ligand binding while the atomic mobility is quite sensitive. Another distinct feature is that the slope of the $\beta = f(\alpha)$ plot is not equal to 2, as expected for a stationary series. At contrary, the slope is much lower, 1.7497 ± 0.1616 which again is in agreement with the observation that the series have a nonstationary character.

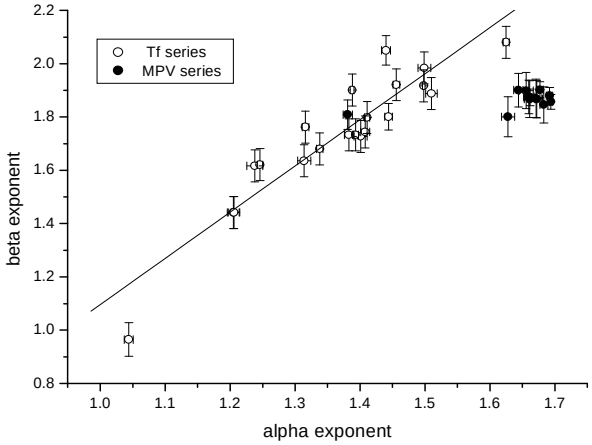

Fig. 2. The scaling exponents α (DFA analysis) and β (spectral analysis) corresponding to the stationary and both nonstationary and stationary respectively contributions for the HIV-1 protease dimer. Each point on the plot corresponds to the same protein with a different bound ligand. The primary data necessary for the calculation of the scaling exponents were collected from the following Protein Data Bank protein entry domains (codes): 1hpv, 1hsg, 1hxw, 1ohr, 7hvp, 1a30, 1axa, 1aid, 1d4s, 1d4y, 1hvr, 1qbs, 1tcx, 1upj, 2aid, 4phv, 6upj, 3upj, 4upj, 1hvh.

While the above analysis of the structure-mobility relationship refers to the long–range properties, a different picture emerges when the structure and mobility series are compared in a direct manner. This is illustrated in Fig. 3 for the same protein.

It is quite evident that there is some kind of relationship between the structure and mobility. The basic resemblance between the curves is that minima or maxima in the two parameters occur at the same atom with practically no exception.

An alternative way to look at the problem is to explore the $T_f = f(MPV)$ function (Fig. 4). This picture does not seem to be very relevant. However if the geometric center of the cell unit of the protein is considered instead of the crystallographic center then a different picture emerges (not shown). In such a case we can noticed a certain regularity of the plot with a minimum in the center of the protein. This aspect will be discussed elsewhere. Therefore a qualitative cross

comparisson of the structure and mobility revels clear connection among them while the nonlinear properties are distinct. In other words using different approaches we can pinpoint both connection and distinction among atomic structure and mobility.

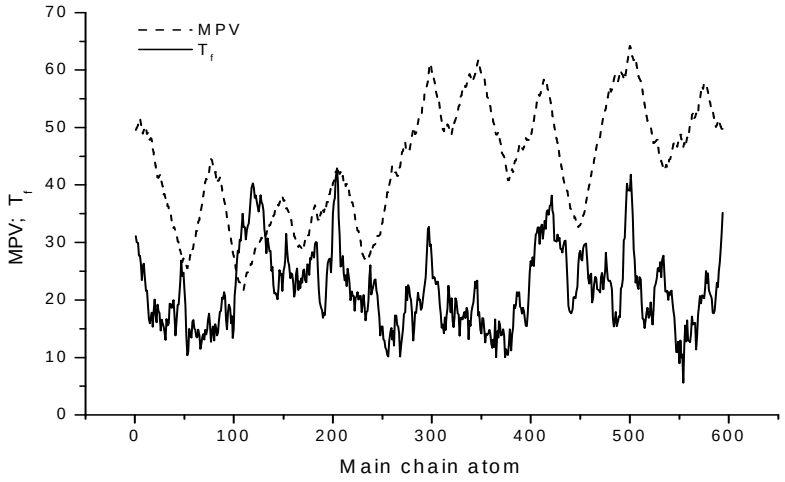

Fig. 3. The Magnitude of the Position Vector, *MPV* and the temperature factor T_f of the protein main chain. The data refer to the HIV-1 protease homodimer (Protein Data Bank code: 1qbs).

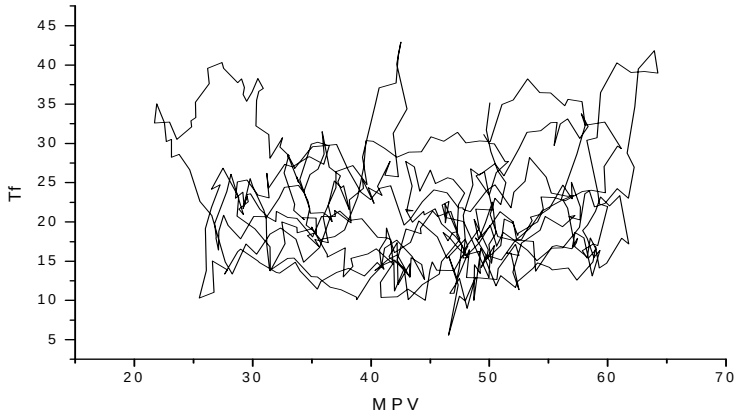

Fig. 4. The relationship between structure and mobility for HIV-1 protease dimer (PDB code: 1qbs) when *MPV* is calculated in the crystallographic center frame.

CONCLUSIONS

The new findings reported by this paper are the following:

a) We report for the first time the long-range correlation properties of the atomic structure in the protein main chains similar to the protein atomic mobility.

b) The long-range correlation structure and mobility behave in a distinct way: while structure is practically insensitive, the mobility is very sensitive to ligand binding. Therefore structure imposes stability on the protein entitiy, while the mobility is related to the specificity of the protein-ligand interaction.

c) The structure seems to have a close to stationary behavior while the mobility have a mixed beahviour consisting of stationary and nonstationary characteristics. It seems that the nonstationary characteristic is the sensitive part to ligand binding.

d) While the long-range correlation properties seem to suggest no direct relationship between structure and mobility, the direct comparisson of the magnitude of the position vector and that of the temperature factor series, at contrary, suggests that there is an evident connection among these two fundamental characteristics of a protein.

Acknowledgements. The work for this paper was financed by a CERES project of the Ministry for Education and Research of Romania.

REFERENCES

1. MORARIU, V.V., A. COZA, Nonlinear properties of the atomic vibrations in protein backbones, *Physica A* , 2003, **320**, 461–474.
2. MORARIU, V.V,, Sationary and nonstationary characteristics of the atomic vibrations in proteins, Proceedings of The 3-rd Conference with international participation "Isotopic and Molecular Processes" Cluj-Napoca, September, 24–27, *Studia Universitatis Babes-Bolyai*, 2003, special issue pp.234–239.
3. PENG, C-K, S. HAVLIN, H.E.STANLEY, A.L. GOLDBERGER, Detrended fluctuation analysis, *Chaos*, 1995, **5**, 82–87.