J. Serb. Chem. Soc. 72 (4) 383–392 (2007)
JSCS–3568

UDC 577.112+547.466:535.337:54.03
Original scientific paper

Analysis of long-range correlation in sequences data of proteins

ADRIANA ISVORAN[1*], LAURA UNIPAN[2], DANA CRACIUN[3] and VASILE MORARIU[4]

[1]West University of Timisoara, Department of Chemistry, Str. Pestalozzi 16, 300115 Timisoara, Romania, [2]University of Agricultural Sciences of Banat, Department of Agriculture, Cl. Aradului 119, 300645 Timisoara, [3]West University of Timisoara, Teacher Training Department, Blvd. V. Parvan 4, 300223 Timisoara and [4]National R&D Institute for Isotopic and Molecular Technology, Department of Molecular and Biomolecular Physics, P. O. Box 700, 400293 Cluj Napoca, Romania (e-mail: aisvoran@cbg.uvt.ro)

(Received 9 November 2005, revised 19 July 2006)

Abstract: The results presented here suggest the existence of correlations in the sequence data of proteins. 32 proteins, both globular and fibrous, both monomeric and polymeric, were analyzed. The primary structures of these proteins were treated as time series. Three spatial series of data for each sequence of a protein were generated from numerical correspondences between each amino acid and a physical property associated with it, *i.e.*, its electric charge, its polar character and its dipole moment. For each series, the spectral coefficient, the scaling exponent and the Hurst coefficient were determined. The values obtained for these coefficients revealed non-randomness in the series of data.

Keywords: sequences of proteins, long-range correlation, spectral coefficient, scaling exponent, Hurst coefficient.

INTRODUCTION

Proteins are polymers or long chains built from a basic set of amino acids. The most common 20 amino acids and their sequence within protein chains are sufficient to create a wide variety of proteins, each suited for its unique function. The amino acids chains are held together by peptide bonds, while particular physical forces existing between the side chains determine the specific spatial structure of a protein.[1] The physical forces involved are:

1) *Electrostatic forces*. Firstly, five amino acids are charged under natural conditions of pH and ionic strength,[1] the two negatively charged amino acids are aspartic acid (Asp) and glutamic acid (Glu) and the three positively charged amino acids are arginine (Arg), lysine (Lys) and histidine (His). This property of amino acid residues determines the electrostatic interactions in proteins.

* Corresponding author.

doi: 10.2298/JSC0704383I

Secondly, some amino acid residues are more electronegative than others and this fact creates local dipoles. The side-chain dipole is one of the strongest contributors to the stability of local structures in proteins. Six of the amino acid side-chains have significant dipole moments,[2] serine (Ser), threonine (Thr), cysteine (Cys), tyrosine (Tyr), asparagine (Asn) and glutamine (Gln).

2) *The hydrophobic forces*. The hydrophobicity of the side-chain is also important for building the spatial structure of proteins. The hydrophobicity is determined by the polarity and size of the amino acid. The less polar a residue is, the more hydrophobic it is. Charged residues and those with strong dipoles are hydrophilic. Neutral residues that do not have polar functional groups are hydrophobic:[2] glycine (Gly), alanine (Ala), valine (Val), leucine (Leu), isoleucine (Ile), methionine (Met), phenilalanine (Phe) and tryptophan (Trp). As a result of the hydrophobic effect, the hydrophobic residues tend to be in the interior of a protein and the hydrophilic residues tend to be on the surface of a protein.

It is very important to know the rules or energy functions involved in this transformation of pieces of one-dimensional information (sequences) into three-dimensional structures (folded proteins).[3] This is not a simple task. The properties of proteins must be considered as complex, heterogenous systems. They can be thought of as disordered systems, but altempts must be made to reveal those non-random features of proteins which are essential for their folding kinetics and native structures.

The aim of this study was to analyze the presence of long-range correlations in sequence data of both globular proteins, enzymes, and fibrous proteins, structural proteins. It is based on treating the protein sequences as time series generated from numerical correspondence between each amino acid and a physical property associated with it: electric charge, dipole moment and polar character. The literature abounds in such type of studies. The nucleotide sequences in deoxyribonucleic acids (DNA) always reveal correlations with different characteristics for the coding and non-coding regions.[4,5] For proteins, a few of them reveal that protein sequences are random and others reveal the contrary, as follows:

i) using the standard run test, White and Jacobs showed that the distribution of hydrophobic residues along sequences was random,[6]

ii) using the non-linear prediction method, Huang and Xiao, on average, did not find significant determistic structures in protein sequences,[7]

iii) using random walk and statistical methods, Pande *et. al*. showed pronounced deviation from pure randomness in the protein sequences, related to the minimization of the energy of the spatial structure,[8]

iv) using spectral analysis, fractal analysis, and statistical thermodynamical tests, Rani and Mitra revealed the presence of regularities on protein sequences,[9]

v) using a correlation function, Weiss and Herzel showed strong correlation related to the hydrophobicity of helix propensity,[10]

vi) there are also many other studies related to the fractal[11–13] and multifractal aspects of the amino acids arrangements.[14]

These conflicting results presented in the literature might be due to inconsistencies in choosing the data sets and algorithms. For example, Pande *et al.*[8] and White and Jacob[6] obtained different results but they used different mapping schemes and different algorithms to test the randomness of the protein sequences. Pande *et al.*[8] considered Lys, His, Arg, Asp and Glu as being hydrophilic amino acids and the others as hydrophobic.[7] In the work of White and Jacob,[6] Phe, Met, Leu, Ile, Val, Cys, Ala, Pro, Gly, Trp and Tyr were considered as hydrophobic and the others as hydrophilic.

It seems that the question whether protein sequences are random, or not, remains open and any new result in this field may contribute to its answer. Here a different mapping scheme from those used by Pande *et al.*[8] and White and Jacob,[6] which corresponded to hydrophobicity, were used. The present mapping schemes were based on the physical properties of each amino acid as presented by Bryngelson and Billings.[2] Also, both statistical and non-linear dynamics methods were employed to analyze the randomness in the series of data. The statistical methods employed were not only the spectral analysis (SA) method,[15] but also the detrended fluctuation analysis (DFA) method,[16] in order to improve the analysis by removing any drift term from fluctuations. The non-linear dynamics method involved the determination of the Hurst coefficient.[17] The presence of long-range correlation in a set of 32 randomly chosen proteins, both globular and fibrous, both monomeric and polymeric polymers were analyzed.

METHOD

The primary structures of the proteins were taken from the Protein Data Bank (http://www.rcsb.org/pdb), the codes entry being:

– 1A3N, 1FAW, 1HBH, 1HDA and 1QPW for haemoglobines belonging to human, bovine, arctic fish, goose and pig, respectively. The multiple sequence alignment for these structures performed using CLUSTALW[18] program showed 62 % mean identity between the primary structures in this case.

– 1LZR, 1HEW, 186L, 1DKJ, 1GD6, 1EL1 and 1BB6 for lysozymes belonging to human, chicken, bacteriophage T4, bobwith quail, *Bombix mori*, *Canis* and trout, respectively. The mean sequences identity in this case was 40 %.

– 1A29, 3CLN, 4CLN and 1CFC for calmodulines belonging to *Bos taurus*, rat, *Drosophila melanogaster* and african frog, respectively. The mean sequences identity in this case was 96 %.

– 1MBA, 5MBN, 2MM1, 1EMY, 1LHS and 1MYT for myoglobins belonging to sea hare, sperm, whale, human, elephant, sea turtle and yellowfin tuna, respectively. The mean sequences identity in this case was 58 %.

– 1BKV for human triple helix of collagen.

– 1NAY for bacterial collagen-like protein.

– 1L9H, 1DZE for bovine and *Heliobacterium rhodopsins*. The sequences identity was 12 %.

– 1GB1 for the B1 domain of *Streptococcus* protein G.

– 1AO6 for human serum albumin and 1ALC for lactalbumin. The sequences identity was 7 %.

– 456C for human metaloprotease.

– 1BR1 and 1N2D for gallus and *Dyctiostelium discoideum* heavy chains of myosin. The sequences identity was 34 %.

The primary sequences of the selected proteins were treated as time series, where the natural succession of the terms in the series may involve a certain correlation. The ideas of the time series were extended to our spatial case of amino acid sequences. The resulting time series were analyzed using methods that enable a correlation in the sequences to be revealed. Three spatial series of data for each sequence of a protein were generated from numerical correspondences between each amino acid and a physical property associated with it. These correspondences generate sequences of numbers. In the present case, the properties were: the electric charge of the amino acid, its polar character, and its dipole moment. The correspondences were as follows:

i) an uncharged amino acid was assigned the number 0, a positively charged one, the number 1 and a negatively charged one, the number –1;

ii) a hydrophilic amino acid was assigned the number 0 and a hydrophobic one, the number 1;

iii) an amino acid with a significant dipole moment was assigned the number 1 and an amino acid with a weak or without a dipole moment, the number 0.

Using the primary structure of the protein and taking into account the physical properties of its amino acids, for each studied protein three series of data were obtained. Then, the following approaches were employed to examine whether there was a long-range correlation within the data set: the SA method,[15] the (DFA) method[16] and the determination of the Hurst coefficient.[17]

The SA method involves the application of a fast Fourier transform to the series under analyzis. This plot gives the power spectrum and for non-linear dynamics it obeys the power law distribution

$$P(f) \approx 1/f^{\beta} \tag{1}$$

where f is the frequency and β is called the spectral coefficient. By representing the power spectrum in a double logarithmical scale, the spectral coefficient is obtained as the slope of the linear fit of the spectrum. As any given time series or spatial series may exhibit a variety of structures, the exponent β ranges in the interval $0 < \beta < 2$. If $\beta = 2$, a power spectral density of white noise is obtained and if $\beta = 1$, a power spectral density of pink noise is obtained. A long-range correlation is present if $0 < \beta < 1$.[17]

The DFA method has also proven useful in revealing the extent of long-range correlation in seemingly irregular time series or spatial series of data. Using the DFA algorithm involves integration the series $u(i)$, $i = 1, 2, \ldots N$:

$$y(i) = \sum_{1}^{i} \left[u(i) - \langle u \rangle \right] \tag{2}$$

where $\langle u \rangle$ is the mean value of $u(i)$. Then the series is divided into boxes of equal size n and the local trend, $y_{\text{local}}(i)$, is calculated for each of the segments by the least squares fit of the data. The local trend is subtracted and the root mean square fluctuation for the given box size is calculated:

$$F(n) = \sqrt{\left[y(i) - y_{\text{local}}(i) \right]^2 / N} \tag{3}$$

If the data are long-range correlated then the relationship between the average fluctuation for a given box, and the box size is of the form:

$$F(n) \approx n^{\alpha} \tag{4}$$

where α is a so-called scaling exponent. The scaling exponent takes the following values: $\alpha = 0.5$ for completely uncorrelated data (white noise), $\alpha = 1$ for $1/f$ noise, $\alpha = 1.5$ for brown noise; $0 < \alpha < 0.5$ for long-range anticorrelation and $0.5 < \alpha < 1$ for long-range correlation.[15] Between the spectral coeficient, β, and the scaling exponent, α, there exists the relation:[19]

$$\beta = 2\alpha - 1 \tag{5}$$

In some cases, the DFA plot has a multislope character. In this work the DFA-1 method, where 1 represents the order of the polynomial function used to fit the local trend, was employed. It reveals both the contribution of the stationary fluctuations and the trend to correlative properties.[20]

Hurst proposed a statistical method to study the time series. For any time series $\{x_k\}_{k=1}^{N}$ and any $2 \le n \le N$, one can define[16]

$$\langle x \rangle_n = (1/n) \sum_{i=1}^{N} x_i \tag{6}$$

$$X(i,n) = \sum_{n=1}^{i} \left(x_n - \langle x \rangle_n \right) \tag{7}$$

$$R(n) = \max_{1 \le i \le n} X(i,n) - \min_{1 \le i \le n} X(i,n) \tag{8}$$

and

$$S(n) = \left[(1/n) \sum_{i} (x_i - \langle x \rangle_n)^2 \right]^{1/2} \tag{9}$$

Hurst found that

$$R(n) / S(n) \propto (n/2)^H \tag{10}$$

where H is called the Hurst coefficient. As n changed from 2 to N, $N–1$ points in the ln*(n) versus* ln*(R(n)/S(n))* plane are obtained. Using the least squares linear fit, the Hurst coefficient is obtained. An exponent greater than 0.5 indicates persistence (past trends tend to persist in the future) whereas an exponent less than 0.5 indicates antipersistence (past trends tend to reverse in the future). An exponent equal to 0.5 indicates randomness in the series of data.[17]

RESULTS

The calculated coefficients for the proteins considered in this study are presented in Table I and their mean values are given in Table II.

TABLE I. The values of the calculated coefficients for the investigated proteins

Protein/property	Electric charge			Hydrophobicity			Dipole moment		
	SA	DFA	Hurst	SA	DFA	Hurst	SA	DFA	Hurst
186L	0.06	0.47	0.47	0.28	0.46	0.86	0.27	0.57	0.76
lHEW	0.09	0.44	0.48	0.26	0.43	0.82	0.14	0.52	0.83
lLZR	0.07	0.43	0.44	0.29	0.46	0.84	0.23	0.53	0.81
lEL1	0.20	0.6	0.54	0.33	0.47	0.86	0.19	0.54	0.84
lBB6	0.15	0.48	0.41	0.40	0.48	0.84	0.26	0.48	0.82
1DKJ	0.09	0.46	0.44	0.24	0.46	0.85	0.03	0.46	0.86
1GD6	0.09	0.54	0.54	0.08	0.45	0.79	0.21	0.47	0.79
1MBA	0.30	0.36	0.38	0.29	0.40	0.89	0.10	0.52	0.72
5MBN	0.06	0.6	0.5	0.31	0.49	0.86	0.10	0.42	0.60
2MM1	0.22	0.5	0.48	0.23	0.52	0.86	0.18	0.52	0.65
1EMY	0.03	0.48	0.46	0.12	0.45	0.82	0.12	0.46	0.71

TABLE I. Continued

Protein/property	Electric charge			Hydrophobicity			Dipole moment		
	SA	DFA	Hurst	SA	DFA	Hurst	SA	DFA	Hurst
1HLS	0.08	0.42	0.44	0.32	0.54	0.84	0.05	0.53	0.68
1MYT	0.10	0.44	0.52	0.40	0.57	0.88	0.36	0.52	0.68
1A3N	0.03	0.44	0.47	0.37	0.43	0.89	0.19	0.55	0.81
1FAW	0.01	0.47	0.48	0.02	0.43	0.89	0.10	0.62	0.84
1HBH	0.05	0.42	0.45	0.15	0.57	0.89	0.19	0.48	0.76
lQPW	0.16	0.42	0.54	0.09	0.45	0.89	0.16	0.53	0.75
lHAD	0.15	0.44	0.46	0.15	0.53	0.91	0.23	0.52	0.76
1N2D	0.05	0.42	0.49	0.18	0.50	0.89	0.08	0.58	0.85
lBRl	0.07	0.47	0.52	0.11	0.44	0.89	0.09	0.57	0.83
lBKV	0.03	0.39	0.69	0.09	0.45	0.85	0.25	0.69	0.65
1NAY	0.56	0.45	0.18	0.03	0.42	0.82	0.03	0.55	0.68
1GB1	0.08	0.4	0.55	0.03	0.55	0.79	0.09	0.56	0.72
lL9H	0.04	0.45	0.45	0.17	0.62	0.91	0.15	0.58	0.81
1DZE	0.08	0.45	0.46	0.16	0.58	0.86	0.18	0.54	0.88
3CLN	0.21	0.57	0.54	0.20	0.42	0.85	0.08	0.49	0.68
4CLN	0.16	0.56	0.57	0.18	0.54	0.84	0.21	0.48	0.75
1A29	0.18	0.57	0.63	0.16	0.53	0.85	0.17	0.54	0.71
1CFC	0.25	0.57	0.63	0.17	0.57	0.83	0.32	0.59	0.76
456C	0.09	0.46	0.43	0.23	0.54	0.85	0.31	0.55	0.77
1A06	0.30	0.46	0.48	0.03	0.52	0.88	0.10	0.46	0.81
lALC	0.04	0.58	0.44	0.10	0.46	0.84	0.13	0.52	0.83

TABLE II. The mean values of the calculated coefficients

Property	Spectral coefficient	Scaling exponent	Hurst coefficient
Electric charge	0.13±0.02	0.47±0.06	0.48±0.08
Hydrophobicity	0.19±0.02	0.48±0.05	0.85±0.03
Dipole moment	0.17±0.01	0.54±0.05	0.76±0.07

The great majority of the values presented in Table I suggests correlations in the data sets for the considered proteins and their mean values clearly indicate it.

Application of the previously mentioned methods are illustrated for sea hare myoglobin. When the SA method is applied, the power spectrum for the electric charges of the amino acids in the sequence of sea hare myoglobin is presented in Fig. 1 on a double logarithmic scale. It can be observed that it shows a power-law distribution. The linear fit of the power spectrum gives the spectral coefficient, β, as the slope of the line. In this case $\beta = 0.30 \pm 0.012$.

Applying the DFA method for the same series, Fig. 2 is obtained. The linear fit of the distribution *F(n) versus n* gives the scaling exponent, α. In this case $\alpha = 0.36$

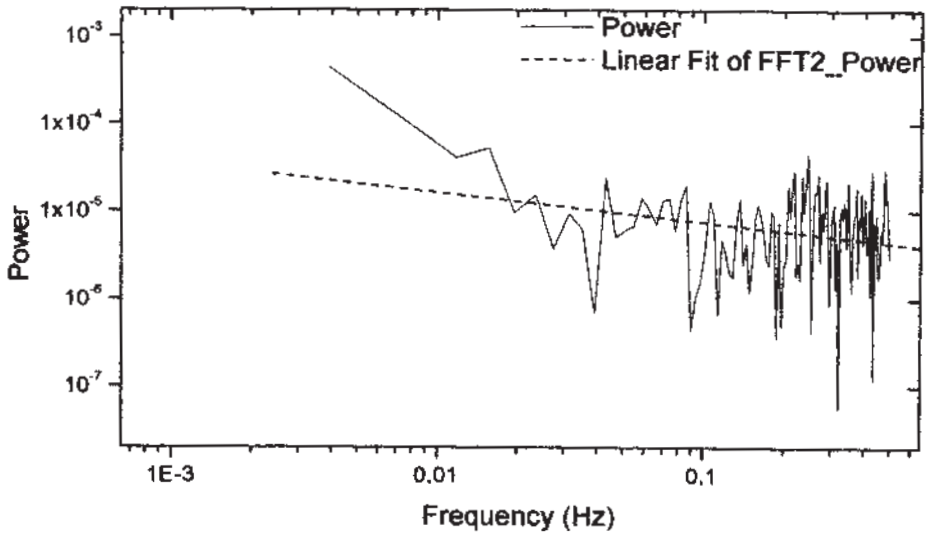

Fig. 1. Determination of the spectral coefficient for the series of electric charges of the amino acid sequence of sea hare myoglobin.

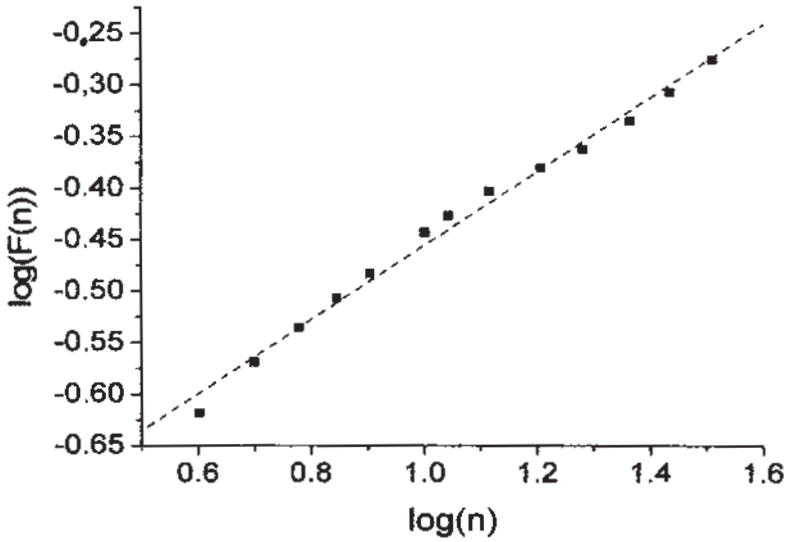

Fig. 2. Determination of the scaling exponent for the series of electric charges of the amino acid sequence of sea hare myoglobin.

± 0.09. The Relationship (5) between the scaling exponent and the spectral coefficient was fulfilled in this case but not for all the series.

For 30 % of the investigated proteins, the DFA plot shows two linear regions. The mean values for the slope of the first region are: 0.38 ± 0.04 for the series obtained using the electric charge, 0.54 ± 0.012 for the series obtained using the hydrophobicity and 0.58 ± 0.07 for the series obtained using the dipole moment.

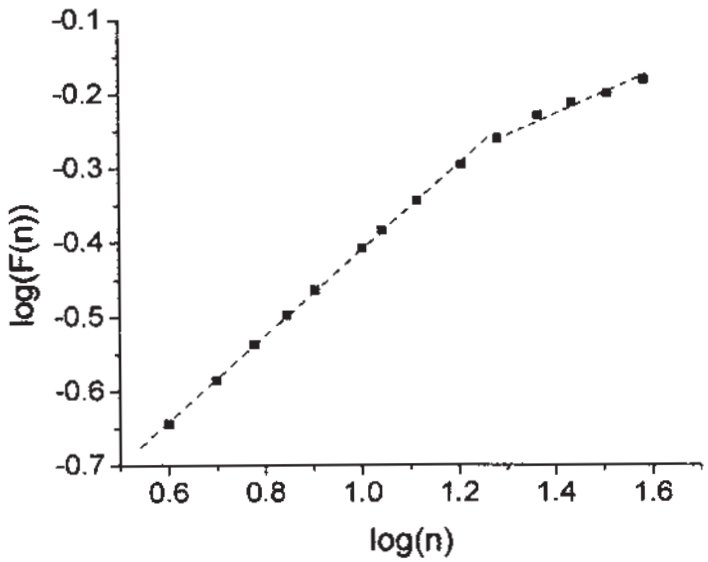

Fig. 3. The DFA plot for the series obtained using the dipole moment for bacteriophage T4 lysozyme.

The value of the slope for the plots obtained using the electric charge indicated anticorrelation and the values of the slope for the series obtained using hydrophobicity and dipole moment indicated weak correlation. The DFA plot for the series obtained using the dipole moments of the amino acids in the primary sequence of bacteriophage T4 lysozyme, is shown in Fig. 3.

The value of the slope for the second region is lower that that for the first region. This indicates a lower correlation at long distances in the series of data. The crossover of the lines occurs at log (n) = 1.22. This means that n = 16.59, which corresponds to the length of the dominant long-range correlation measured here for a number of amino acids. A similar result was obtained for the correlation in a series of temperature factors of atoms belonging to the backbones of 50 randomly chosen proteins.[21]

DISCUSSION AND CONCLUSIONS

Proteins are informational molecules and their primary sequences hold the key to their tertiary structure and to their biological activity. The tertiary structure is folded and it can be extended, semi-compact or compact.[1] Protein folding can bring together some amino acid residues which may be physically distant in the primary structure and such a correlation is ussually very important for retaining the folded form.[1] The role of long-range interactions of the residues in defining the secondary structure of a protein has already been demonstrated.[23]

Each protein can be considered as arising from a dispersion of amino acids in the sequence space. Each amino acid begins to search its own local region of the sequence space using physical criteria.[1] Here, the electric charge of the amino acid,

its polarity and its dipole moment, were considered as criteria. There are three basic types of behavior that might be expected. The investigated sequences can be random, they might indicate a persistent or an antipersistent behavior. A random trajectory expresses that the sequence has no memory.

Almost all the values for the calculated coefficients indicate correlation in the sequences of proteins (see Table I). The values obtained for all the Hurst coefficients clearly reveal correlations in the data sets. The values of the spectral coefficients and scaling exponents do not indicate this very clearly, some of their values are closer to those corresponding to random series. However, the main difficulty when such studies are employed is that the transformation of the amino acid arrangement into numerical sequences based on different physical properties does not conserve all the statistical properties of the initial sequence. Thus, Fourier transformation might miss some periodicities of the sequence. There are other examples in the literature where the SA technique does not clearly reveal correlation, although other techniques did.[22] As mentioned before, some of the DFA plots had a multi-slope character and the non-stationarity in the series affects the values of the scaling exponents. The mean value of the spectral coefficient with respect to that of the scaling exponent indicates correlation in the data sets.

In the cases of the series obtained using the dipole moment and the hydrophobicity of the amino acids, the values of the Hurst coefficients indicate persistence (see Table I). This was also the case for the mean values of the scaling exponents of the series obtained using the dipole moment. For the series obtained using the hydrophobicity of the amino acids, the mean value of the scaling exponents was very close to that indicating randomness of the series, but this may be the result of non-stationary fluctuations in many series of data. In a persistent behavior, the presence of one characteristic type of amino acid in the primary structure of a protein increase the probability of the further appearance of an amino acid with the same characteristic. The results are in good agreement with the experiments and simulation data, which indicate the presence of packed hydrophobic regions in the interior of proteins and of hydrophilic regions exposed to the solvent. Also to be expected are regions with dipolar amino acids in the protein structure.

In case of the series obtained using the electric charge as the characteristic property of an amino acid, the mean values of the Hurst coefficients and the scaling coefficients indicate a weak antipersistence. An antipersistent behavior means that the appearance of an amino acid with a characteristic property increases the probability of the appearance of an amino acid with the opposite property. This result is also not surprising because proteins require favorable electrostatic interactions to build their tertiary structures.

The results presented here indicate correlation and non-randomness in the sequences of the investigated proteins. This result is important becaus it may be related to the laws of protein folding and may contribute to a better understanding of the processes which enable a protein to perform its function.

ИЗВОД

АНАЛИЗА КОРЕЛАЦИЈЕ ДУГОГ ДОМЕТА У ПОДАЦИМА О СЕКВЕНЦИЈИ ПРОТЕИНА

ADRIANA ISVORAN,[1] LAURA UNIPAN,[2] DANA CRACIUN[3] и VASILE MORARIU[4]

[1]West University of Timisoara, Department of Chemistry, Str. Pestalozzi 16, 300115, Timisoara Romania, [2]University of Agricultural Sciences of Banat, Department of Agriculture, Cl. Aradului 119 300645 Timisoara, [3]West University of Timisoara, Teacher Training Department, Blvd. V. Parvan 4, 300223 Timisoara i [4]National R&D Institute for Isotopic and Molecular Technology, Department of Molecular и Biomolecular Physics, P. O. Box 700, 400293 Cluj Napoca, Romania

Приказани резултати сугеришу постојање корелација у подацима о секвенцији протеина. Анализирана су 32 протеина, како глобуларна тако и фибрилна, како мономерна тако и полимерна. Примарна структура протеина третирана је као временски успостављени низ. За сваку секвенцију протеина генерисана су три просторна низа података коришћењем повезаности између сваке амино киселине и физичке величине (наелектрисање, поларни карактер и диполни момент) која јој одговара. За сваки низ одређен је спектрални коефицијент, експонент скалирања и Hurst-ов коефицијент. Вредности добијене за ове коефицијенте показале су да нема случајности унутар низова података.

(Примљено 9. новембра 2005, ревидирано 19. јула 2006)

REFERENCES

1. D. Voet, J. G. Voet, C. W. Pratt, *Fundamentals of Biochemistry*, J. Wiley, New York, 2002, p. 77
2. J. D. Bryngelson, E. M. Billings, in *Physics of Biologic Systems,* H. Flybjerg, Ed., Springer, Berlin, 1997, p. 23
3. P. G. Walymes, in *Physics of Biologic Systems,* H. Flybjerg, Ed., Springer, Berlin, 1997, p. 56
4. M. Martinis, in *Supramolecular Structures and Functions*, G. Pifat-Mrzljac, Ed., Kluwer Academic/ Plenum Publishers, New York, 2001, p. 185
5. Y. Zu-Guo, A. Vo. L. Ka-Sing, *Phys. Rev. E.* **64** (2001) 031903
6. S. T. White, R. E. Jacobs, *Biophys*. *J.* **57** (1990) 911
7. Y. Huang, X. Yi, *Chaos Solutions Fractals* **17** (2003) 895
8. V. S. Pande, A. Y. Grosberg, T. Tanaka, *Proc. Nat. Acad. Sci. USA* **91** (1994) 12972
9. M. Rani, C. K. Mitra, *J. Biomol. Struct. Dyn.* **13** (1996) 935
10. O. Weiss, H. Herzel, *J. Theor. Biol.* **4** (1998) 341
11. T. G. Dewey, *Fractals in Molecular Biophysics,* Oxford University Press, New York, 1997, p. 72
12. C. X. Wang, F. H. Huang, *Phys. Rev. A* **41** (1990) 7043
13. A. Isvoran, A. Licz, L. Unipan, V. V. Morariu, *Chaos Solitons Fractals* **12** (2001) 757
14. A. Isvoran, *Chaos Solitons Fractals.* **19** (2004) 141
15. P. Szendro, G. Vincze, A. Szasz, *Eur. Biophys. J.* **30** (2001) 227
16. C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger, *Phys. Rev. E* **49** (1994) 1685
17. J. C. Sprott, G. Rowlands, *Physics Academic Software*, American Institute of Physics, 1995
18. C. Ramu, S. Hideaki, K. Tadashi, L. Rodrigo, G. Toby, D. Higgins, J. Thompson, *Nucleic Acids Res.* **31** (2003) 3497
19. K.Hu, P. C. Ivanov, Z. Chen, P. Carpena, H. E. Stanley, *Phys. Rev. E.* **64** (2001) 011114
20. J. Argyris, G. Faust, M. Haase, *An Exploration of Chaos,* North-Holland, Amsterdam, 1994, p. 124
21. Y. Zu-Guo, B. Wang, *Chaos Solitons Fractals* **12** (2001) 519
22. M. A. Korotkova, E. V. Korotkov, V. M. Rudenko, *J. Mol. Mod.* **5** (1999) 103
23. D. Kihara, *Protein Sci.* **14** (2005) 1955.