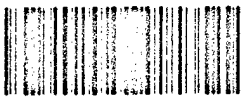


PB93-210961

Amino Acid Sequence Transfer Operators and Metric Space Distortions of Proteins

Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette (France)

Prepared for:

Office of Naval Research, Arlington, VA

May 93

BIBLIOGRAPHIC INFORMATION

PB93-210961

Report Nos: IHES/M-93/22

Title: Amino Acid Sequence Transfer Operators and Metric Space Distortions of Proteins.

Date: May 93

Authors: A. J. Mandell, and K. A. Selz.

Performing Organization: Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette (France).**Florida Atlantic Univ., Boca Raton.

Sponsoring Organization: *Office of Naval Research, Arlington, VA.

Supplementary Notes: Prepared in cooperation with Florida Atlantic Univ., Boca Raton. Sponsored by Office of Naval Research, Arlington, VA.

NTIS Field/Group Codes: 57F, 57B, 99F

Price: PC A02/MF A01

Availability: Available from the National Technical Information Service, Springfield, VA. 22161

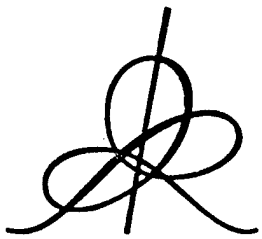
Number of Pages: 6p

Keywords: *Amino acid sequence, *Proteins, *Protein conformation, *Enzymes, X ray diffraction, Crystals, Thermodynamics, Riemannian manifolds, Hydrophobicity.

Abstract: Physical and chemical studies of protein dynamics often yield nonlinear distortions and multiplicities of time scales which are partially resolved via scaling solutions. Assuming that the internal basis of the metric relations of proteins is hydrophobic hydration, we demonstrate a relationship between an expanding group action on hydrophobic amino acid sequences and the fractional spatial scalings of amino acid monomeric mass distances in 25 representative proteins. This empirical relationship is analogous to that resulting from the tessellation of non-Euclidean spaces by discrete groups and suggests a route to physical uniformization of the data of protein dynamics without ad hoc scaling corrections.

AMINO ACID SEQUENCE TRANSFER OPERATORS AND METRIC SPACE DISTORTIONS OF PROTEINS

A. J. MANDELL K.A. SELZ



Institut des Hautes Etudes Scientifiques
35, route de Chartres
91440 - Bures-sur-Yvette (France)

Mai 1993

IHES/M/93/22

AMINO ACID SEQUENCE TRANSFER OPERATORS AND METRIC SPACE DISTORTIONS OF PROTEINS

Arnold J. Mandell and Karen A. Selz
Department of Mathematics, Physics, Psychology
and the Clinical Psychology Center
Florida Atlantic University, Boca Raton, FL33431

(Current Address: Institut des Hautes Etudes Scientifiques)

April, 1993

Physical and chemical studies of protein dynamics often yield nonlinear distortions and multiplicities of time scales which are partially resolved via scaling solutions. Assuming that the internal basis of the metric relations of proteins is hydrophobic hydration, we demonstrate a relationship between an expanding group action on hydrophobic amino acid sequences and the fractional spatial scalings of amino acid monomeric mass distances in 25 representative proteins. This empirical relationship is analogous to that resulting from the tessellation of non-Euclidean spaces by discrete groups and suggests a route to physical uniformization of the data of protein dynamics without ad hoc scaling corrections.

Diffraction analyses on classical crystals yield reciprocal space patterns from Fourier transformations of periodic delta functions with translational invariance[1]. In contrast, aperiodic, less regular patterns, characteristic of many proteins, support symmetry operations which involve less an invariance against translation and more against changes in scale,[2-5]which we have suggested may reflect distortions in the metric space of hydrophobic hydration[6]. Protein crystals range from 30% to 78% in solvent content[7]. The spatial distortion idea might be best understood in the context of a simple analogy.

A space is non-Euclidean when the distance, d, between two points, x and y, is no longer given by the theorem of Pythagoras, $d = \sqrt{x^2 + y^2}$, thus requiring a "metric tensor," $\phi(l)$ to establish the length spacings of ticks on its ruler. Negatively curved space "crinkles" when displayed in the Euclidean metric such that moving equal sized steps, $l_i = 1, 2, 3, \ldots, 8, 9, 10$ in curved space (use a as the negative curvature parameter) would transform as $\phi(l) = l' = \frac{l+a}{1+al}$ in Euclidean space, if $a = .2, l_i = 1, 1.57, 2.0, \ldots, 3.15, 3.29, 3.4$; if $a = .3$, then in Euclidean space, the $l_i$ticks would occur at $= 1, 1.44, 1.74, \ldots, 2.44, 2.51, 2.58$. Making measures of length of an object with an intrinsically curved metric from a Euclidean point of view would, of course, require the rescaling of length, $\phi(l)^{-1} = l = \frac{l'-a}{1-al'}$, in order to obtain its real spatial invariants.

This report of a relationship between a protein packing scaling number, S, and measure, h, made on amino acid hydrophobic sequence transfer functions, $\mathcal{L}$, in a non-Euclidean setting is cast as a problem analogous to one relating geometric packing laws to the actions of fundamental groups characterizing the space of the packings. An anatomy of globular proteins in curved space is suggested by the fact that regular close packings that also satisfy the tetrahedral hydrophobic bonding requirements of protein components in water are not possible in Euclidean space, but may be so as tetrahedral joinings in hyperbolic space[8,9]. The solvent mediated, enthalpy-entropy compensation mechanisms of organic chemistry[10]and proteins in water[11]demonstrate isoenergetic rearrangements of hydrogen bonds to satisfy constant density and minimal surface demands of hydrophobic cavities[12] creating clathrate coordination numbers of a great variety[13]. Distortions in hydrophobic hydration space by singular, dense, hydrophobic "knots"[14]in the interiors of globular proteins have been postulated to be crucial for their tertiary structures in water. Curvature arises from these kinds of "laws of bonding." For example, in the plane where the coordinate number of a triangular tessellation is six, if the coordination number is less than six, the space must be curved positively and if greater than six, a two dimensional space of negative curvature results.

Using the X-ray coordinates of α-carbons to establish a range of inter-α-carbon distances, a series of studies[15,16]indicated that there were replicable fractional power laws relating the amino acid monomeric mass density and the α-carbon distance metric which we call the Stapleton Number, S, after its discoverer. S values ranged from "curled up" α-helical, barrel dominated proteins such as hemoglobin-α(equine) and myoglobin (sperm whale) of $S = 1.65$ and 1.67, respectively, to "extended," more random protein conformations such

as those of protease A and protease B(s. griseus) for which $S = 1.31$ and 1.30. Of interest with respect to the functional meaning of S is that $S = 1.31$ for both soy bean trypsin and trypsin inhibitor. We regard S as a crude index of the relative putative curvature of the protein's natural metric space.

The amino acid sequence transfer operator, $\mathcal{L}$, for each protein was constructed as follows: each of the twenty amino acids were assigned to one of four hydrophobic equivalence classes, $\{A, B, C, D\}$, consistent with typical results of water-organic solvent equilibrium concentration studies.[17,18] $\{A, B, C, D\}$ represents a non-arbitrary, natural partition.[19] The assignments were: A = {SER, THR, GLY, GLN, ASN}; B = {ALA, ASP, HIS, ARG, GLU}; C = {CYS, MET, VAL, LYS, LEU}; D = {TRP, TYR, PHE, PRO, ILE}.

The SER, ILE, VAL, ALA, ARG ..., peptide sequence in symbolic dynamics would read: $\mathcal{L} : A \rightarrow D \rightarrow C \rightarrow B \rightarrow B \ldots$. The protein sequence was be encoded and counted via 4×4 transition matrices, $\mathcal{L} : M_{i,j} \rightarrow M_{i,j}$, $\mathcal{L} : \{A, B, C, D\} \rightarrow \{A, B, C, D\}$, via the actions of the transfer operator, $\mathcal{L}$, working its way down the protein amino acid hydrophobic sequences. $M_{i,j}$ was then transformed into a "sparce" incidence matrix, $B_{i,j}$, by encoding each $M_{i,j}$ as 1 if its contents were equal to or greater than 4% of the total number of amino acids in the protein and 0 if it were below it.

In negatively curved space, straight lines, called geodesics, are distorted into curved loops[21]. The measure, h, on the number of loops (in hydrophobic symbols) of length less than or equal to n under the action of $\mathcal{L}$ can be estimated generally as $h(\mathcal{L}) = \frac{\log(\text{trace}(B^n))}{n}$ which converges with n[21]. The use of a similar counting functions for geodesics of conjugacy classes of fundamental groups of hyperbolic surfaces is common in the ergodic, measure theoretic approach to dynamical systems on negatively curved (expanding) manifolds (surfaces), and is often called the Ruelle-Frobenius-Perron zeta function[22]. In our empirical study, greater curvature indices, S, should be associated with larger values of $h(\mathcal{L})$. For our initial examples, hemoglobin and myoglobin, $h(\mathcal{L}) = 0.82$ and 1.01; while for protease A and B, $h(\ell) = 0.69$ and 0.48 respectively.

The proteins studied were: protease B, protease A, (*s. griseus*), superoxide dismutase (*bovine*); elastase (*porcine*), subtilisin inhibitor (*s. albogriscoulus*), algal ferredoxin (*s. platensis*), trypsin inhibitor (*bovine*), carbonic anhydrase B(*human*), dihydrofolate reductase (*l. casei*), staphylococcal nuclease (*s. aureus*), flavodoxin (*clostridium*), alcohol dehydrogenase (*equine*), glyceraldehyde dehydrogenase (*lobster*), cytochrome $B5$(bovine), rhodanese (*bovine*), carboxypeptidase A (*bovine*), cytochrome $C2$(*r. rubrum*), lactate dehydrogenase (*dogfish*), hemoglobin-β(*equine*), adenylate kinase (*porcine*), thioredoxin (*e. coli*), hemerythrin (*t. dyscritum*), myoglobin (*sperm whale*), lysozyme (*bacteriophage T4*), and thermolysin (*b. thermoproteolyticus*).

A linear regression yielded $S = 0.881 + 0.710(h(\mathcal{L}))$ for the 25 representative proteins and a strong linear correlation coefficient between S and $h(\ell)$ was indicated by a correlation coefficient $r = 0.654$. Since the average amino acid hydrophobicity, (kcal K^{-1}mol^{-1})[18], as well as the density of "loops," $h(\mathcal{L})$, contribute to water space distortion as S, a future study will involve a model which will incorporate variables relating "magnitudes and frequences" of hydrophobic amino acids in the protein sequences to S. Our findings are qualitatively consistent with the idea that amino acid hydrophobicity sequence transfer functions can serve as statistically descriptive predictors of the characteristic geometry of their macromolecules hydrophobic hydration space.

Edsall[23] was the first to attribute the anomalous increases in heat capacity of nonpolar solutes in water to the formation of "structured water." Chothia[24], using x-ray crystallographic and hydrophobicity data for amino acid side changes described anomalously small volumetric changes associated with protein (lysozyme) unfolding which from our point of view already implicated a "crinkled-in-Euclidean," curved metric of the water space. Minimal surfaces in theory and soap-like substances, such as glyceryl mono-oleate, pack surfaces of negative curvature[9]. The implications of a natural, non-Euclidean metric for the apparent nonlinear distortions and multiplicities of time and space scales in physical and chemical measurements made on protein dynamics are obvious[2-6,25-28].

REFERENCES

1. Blundell, T.L. and Johnson, L.N. *Protein Crystallography* (Academic Press, New York, 1976).

2. Frauenfelder, H.G., Petsko, G.A. and Tsernoglou, D. *Nature* 280, 558 – 563(1979).

3. Stapleton, H.J., Allen, J.P., Flynn, C.P., Stinson, D.G., and Kurtz, S.R. Phys. Rev. Lett. **45**, 1456 -- 1459(1980).

4. Stein, D.L. *Proc. Natl. Acad. Sci. USA* **82**, 3670 – 3674(1985).

5. Elber, R. and Karplus, M. *Science* **235**, 318 – 319(1987).

6. Mandell, A.J., Russo, P.V., and Blomgren, B.W. *Ann. N. Y. Acad. Sci.* **504**, 88 – 117(1987).

7. McPherson, A. *Preparation and Analysis of Protein Crystals* (Wiley, New York, 1982).

8. Coxeter, H.S.M. *Proc. Roy. Soc.* **A278**, 147 – 167(1964).

9. Mackay, A.L. *Physica* **113B**, 300 – 305(1985).

10. Leffler, J.E. *J. Org. Chem.* **20**, 1202 – 1231(1955).

11. Lumry, R. and Rajender, S. *Biopolymers* **9**, 1125 – 1227(1970).

12. Kuntz, I.D. and Kauzmann, W. *Adv. Protein Chem.* **28**, 239 – 345(1974).

13. Stillinger, F.H. *Science* **209**, 151 – 457(1980).

14. Lumry, R. and Gregory, R.B. in: *The Fluctuating Enzyme* (ed. Welch, G.R.)1 – 190(Wiley, New York, 1986).

15. Allen, J.P., Colvin, J.T., Stinson, D.G., Flynn, C.P. and Stapleton, H.J. *Biophys. J.* **38**, 299 – 310(1982).

16. Wagner, G.C., Colvin, J.T., Allen, J.P., and Stapleton, H J. *J. Am. Chem. Soc.* **107**, 5589 – 5594(1985).

17. Nozaki, Y. and Tanford, C. *J. Biol. Chem.* **246**, 2211 – 2217(1971).

18. Manavalan, P. and Ponnuswamy, P.K. *Nature* **275**, 673 – 674(1978).

19. Paulus, M.P., Geyer, M.A., Gold, L.J. and Mandell, A.J. *Proc. Natl. Acad. Sci. USA* **87**, 723–727(1990).

20. Bearden, A.F. *The Geometry of Discrete Groups* (Springer, New York, 1983).

21. Margulis, G. *Functional Anal. Appl.* **3**, 89 – 90(1969).

22. Ruelle, D. *Invent. Math.* **34**, 231 – 242(1976).

23. Edsall, J.T. *J. Am. Chem. Soc.* **57**, 1506 – 1507(1935).

24. Chothia, C. *Nature* **254**, 304 – 308(1975).

25. Careri, G., Fasella, P. and Gratton, E. *CRC Crit. Rev. Biochem.* **3**, 141 – 164(1975).

26. Knox, D.G. and Rosenberg, A. *Biopolymers* **19**, 1049 – 1057(1980).

27. Liebovitch, L.S. and Toth, T.I. *Ann. N. Y. Acad. Sci.* **591**, 375 – 391(1990).

28. Tian, W.D., Sage, J.T., Srajer, V., Champion, P.M. *Phys. Rev. Lett.* **68**, 408 – 411(1992).

ACKNOWLEDGEMENTS. This work was supported by the (USA) Office of Naval Research (Theoretical Biophysics and Biological Intelligence). Appreciation is expressed to the Institute des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France, for their hospitality during the theoretical development and writing of this paper.