Tel Aviv University
Raymond and Beverly Sackler Faculty of Exact Sciences
School of Chemistry

The Fractal-Like Nature of Proteins:

Foundations, Applications and Ramifications

Shlomi Reuveni

Submitted as a Ph.D Thesis in Chemistry

March 2012

This work was carried out

under the supervision of

Prof. Joseph Klafter

School of Chemistry, Tel Aviv University

and the guidance of

Prof. Rony Granek

Department of Biotechnology Engineering,
Ben-Gurion University, Beer Sheva

Contents

1 Acknowledgments

This thesis, a product of five industrious years, concludes a joyful period I have spent as a Ph.D. student in the school of chemistry at Tel-Aviv University. Looking back in retrospect on the amazing time I have had, I know that I couldn't have possibly asked for more. Although I feel lucky, I know that my good fortune is not a matter of sheer chance. And so, before we start, I would like to express my deep gratitude to a few special people that made all of this possible.

First and foremost, I feel extremely fortunate to have had the wonderful opportunity of studying under the supervision of professor Yossi Klafter. I consider Yossi a mentor for life, a nobleman and a mensch. In his unique and humble way Yossi taught me as much about science as he did about scientists and human beings in general. His support in me and his genuine care and warmth were never ending and I will forever cherish them in my heart.

I am also in great debt to professor Rony Granek, a truly great physicist whose depth and intellect never cease to amaze me. Rony was there for me, day and night, to guide and nurture. Working by his side, I have learned numerous lessons and my skills as a physicist have evolved considerably. Rony's wisdom, modesty and intellectual integrity are exemplary and I can only try and follow the path he has paved.

Writing this thesis could not have been possible without the generous financial support of the Israeli council for higher education by means of its 'converging technologies' program. I thus thank them with all my heart and pray, for the sake of others, that similar initiatives will take place in the future as well. Financial and moral support from the 'interdisciplinary program for outstanding students', named after Adi Lautman, were not less crucial. The interdisciplinary program is an enchanted heaven, studying there has changed my life and for that I will be grateful forever.

In a more personal note, I would like to thank my family for their unconditional support in me. Mom, Dad, Rinat, although I will never be able to operate, hopefully you would soon be able to say you have a 'doctor' in the family! Last but not least, I would like to thank Shira, my partner for life, who is the only person other than me, who really knows about the times where this was no picnic. Shira, I love you.

2 Thesis Outline

This thesis is a compilation of articles revolving around a single general theme - *the fractal-like nature of proteins*. The conjecture that fractals may have something to do with proteins is not new. However, in the last few years there has been renewed interest in the fractal-like nature of proteins as studies have pointed out numerous similarities between proteins and fractals. In particular, proteins were shown to resemble fractals in the way in which they thermally vibrate and in the manner in which they fill space. Motivated by these observations we have set off to harness the vast mathematical and physical machinery, originally devised to describe the physical properties of fractals, in order to quantitatively analyze protein structure and dynamics.

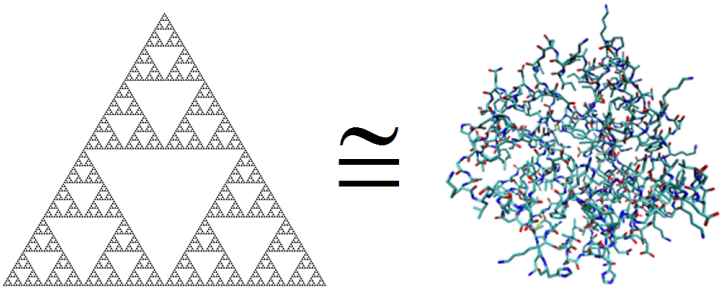


Figure 1: Proteins are large organic molecules that play a vital role in all biological organisms. Fractals are geometrical objects that possess self-similarity. The Sierpinski gasket fractal (left) side by side with a 3D visualization of the protein triose phosphate isomerase (right). The physics of fractal systems can shed new light on the vibrational dynamics of proteins.

The heart of this manuscript is a series of seven articles which, after being reviewed, appear fully in chapter 9. It is important to stress that all the articles presented herein are connected by a continuous line of reasoning and together they form a single body of work which is inseparable.

Chapters 4-8 precede chapter 9 and are intended to introduce the reader to the research topic and the basic mathematical and physical tools we have utilized and advanced upon at a latter stage. Most of the material that appears in these chapters can be found scattered throughout the literature. I have nevertheless thought it would be advantageous to gather relevant information and present it here while giving special attention to the context of this work. Readers already proficient in the subjects discussed below can skip the introduction and start reading directly from chapter 9.

Chapter 4 provides a short introduction on proteins and fractals. A special emphasis is put on structure and on the connection between structure and function. The fractal dimension is introduced and calculated for both fractals and proteins, a first link between the two is formed. From structure we move on to dynamics which is discussed in chapter 5. After briefly reviewing molecular dy-

namics simulations and Go-like models we continue to describe elastic network models and the Gaussian Network Model (GNM) in particular.

Most analytical treatments of protein dynamics entail a compromise between physical realism and mathematical tractability. The challenge is to identify a simple, yet physically plausible, model that retains at least some properties of interest and experimental relevance. To that extent the GNM has served us well and we have used it extensively in all stages of our research. By discussing some mechanistic aspects of the GNM we set the stage for the introduction of the spectral dimension which plays a prominent role in what follows. We define the spectral dimension in the case of general fractal networks and point out its existence for proteins. An additional link between proteins and fractals is forged.

Chapters 6 and 7 discuss thermal vibrations within the framework of the GNM. We start with static (time independent) quantities and focus on the static vibrational mean square displacement. The concept of Landau-Peierls instability is introduced and an instability criterion is derived. Implications on the stability of fractals and proteins are discussed as an appetizer for what will follow. We continue with dynamic quantities and review the Langevin equation together with some classical results regarding Langevin dynamics. The Langevin equation for an elastic network coupled to a thermal bath is taken as a starting point in a few of our articles. We thus take the time to derive the Langevin equation for this special case and do so in both regular and normal coordinates.

After discussing the Langevin equation we continue to chapter 8 which concludes the introduction. In this chapter we discuss random walks and the mapping between random walks and thermal vibrations. Our interest in random walks (in the context of this thesis) started when we have realized that they can be used as a practical tool in the study of protein structure and dynamics. Later on, we have advanced on some theoretical aspects of the mapping between random walks and thermal vibrations and so, a brief introduction to this intriguing topic is most definitely in place.

In our studies we have tried to examine the boundaries of the analogy between proteins and fractals and often asked ourselves which part of what can be said about fractals remains true for proteins as well? And yet, as is often the case in scientific research, our work was also prone to unexpected ramifications. We could not have guessed that while trying to gain some additional insight about the vibrational dynamics of proteins we will end up discovering new relations between random walks and thermal vibrations. Surely, one could not have foreseen a connection between the solution of the first passage time problem, in complex scale invariant media, and the variance in the distance between two fluctuating amino acids. Applications of our findings to the understanding of structures as diverse as glass forming colloidal suspensions and chromatin were also quite unexpected and yet, this is what science is all about. The thesis is sealed in chapter 10 where we recapitulate the main results we have obtained and highlight important implications and applications.

3 Hebrew Summary

חיבור זה הינו אסופת מאמרים המוקדשת לחלבונים, לקשרים בינם ובין אובייקטים מתמטיים הנקראים פרקטלים וכן למשמעויות, ההשלכות והיישומים של קשרים אלו. פרקים ארבע עד שמונה מהווים את ההקדמה והרקע הכללי לעבודה ופרק עשר מוקדש לסיכום ומסקנות. את הלב של חיבור זה ניתן למצוא בפרק תשע, הסוקר בהרחבה את המאמרים, תוך שימת דגש על הקשר הלוגי ביניהם, כפי שמשתקף מהשתלשלות המחקר. לאחר סקירה זו, מובאים המאמרים בשלמותם כפי שהתפרסמו בעיתונות המדעית. תקציר העבודה ועיקריה מובאים כאן בשפה העברית.

חלבונים הם מולקולות אורגניות גדולות המסונתזות באופן טבעי ע"י יצורים חיים. חלבונים נמצאים בתאיו של כל יצור חי, ללא יוצא מן הכלל, ובתאים עצמם נמצאים החלבונים כמעט בכל מבנה ואברון שהוא. מכלול תכונותיו של האורגניזם (הפנוטיפ) - צבע עיניים, פעילות מערכת החיסון, רגישות לחומרים מסוימים וכו' - הינו תוצאה ישירה של פעילות חלבונים.

פרקטלים הם אובייקטים מתמטיים הניחנים בתכונות מעניינות רבות. אחת מהתכונות הללו מכונה דמיון עצמי, והיא מבטאת את העובדה שאם נבחן חלק קטן מפרקטל ידמה חלק זה לשלם כולו. חלבונים אינם ניחנים בכל התכונות המאפיינות פרקטלים, לפחות לא באותה מידה של אידיאליות, אך מסתבר שניתן בכל זאת להתבונן בחלבונים מבעד למשקפיים פרקטליות וליחס מאפיינים פרקטלים גם להם.

התיאור הפרקטלי של חלבונים אפשר לנו לבחון באור חדש הן את המבנים התלת ממדיים האופייניים למולקולות חלבון והן את הדינמיקה המאפיינת את האופן שבו מולקולות אלו רוטטות בתגובה להתנגשויות עם מולקולות הממס האופפות אותן. בדומה לפרקטלים ניתן לייחס לכל חלבון שני ממדים האופייניים לו. מניתוח המבנה הגבישי של חלבון ניתן להסיק את המימד הפרקטלי ומניתוח מודל דינמי של חלבון ניתן להסיק את המימד הספקטראלי. ממדים אלו מהווים מעין תזקיק של התכונות המבניות והדינמיות, בהתאמה, של אותו חלבון. הידע הרב שנצבר במהלך ניתוח התכונות הפיסיקליות והמתמטיות של פרקטלים שימש אותנו בבואנו להתבונן בחלבונים מזווית חדשנית ושונה זו. מחקרנו שילב שימוש במודלים תאורטיים העוסקים בפיזיקה של מערכות פרקטליות אלסטיות עם עבודה ביואינפורמטית שבוצעה תוך שימוש במספר מאגרי מידע קיימים ותוך יצירת בסיס נתונים חדש.

חלבונים מתאפיינים באיזון עדיין בין יציבות וגמישות. על מנת לבצע את תפקידם כהלכה עליהם לשמור על מבנה תלת ממדי מוגדר, שכן קיים קשר הדוק בין מבנה לתפקוד. מאידך, גמישות מבנית מסוימת דרושה אף היא לצורך תפקוד תקין. עקרון אי היציבות של לנדאו ופיירלס מאפשר לכמת את האיזון העדיין בין יציבות וגמישות, כך שניתן יהיה להביעו כ-'משוואת מצב, החוזה את הקשר בין מספר חומצות האמינו המרכיבות חלבון לממדים הפרקטלים המאפיינים אותו [47]. את התחזית העולה ממשואה זו בדקנו בהצלחה אל מול הניסיון במחקר שכלל כחמש מאות חלבונים שונים [47]. המסקנה העיקרית העולה מהמחקר היא שכאשר בוחנים את התכונות הפרקטליות של חלבונים, ניתן למצוא קווים ברורים של דמיון ביניהם, וזאת למרות המגוון הקיים. דומה כי זכוכית המגדלת הפרקטלית חושפת קשרים חבויים בין מבנה ודינמיקה וכי חלבונים השונים זה מזה תכלית שינוי, מצייתים כולם למשוואה אוניברסלית אחת, המקשרת בין אורך דינמיקה וטופולוגיה. כחלק ממחקר המשך, הורחב שדה הנתונים בצורה משמעותית ונבדקו יותר מחמשת אלפים חלבונים. תוצאות המחקר המורחב איששו את המסקנות שעלו מן המחקר המקורי [48].

המשך המחקר הוקדש לפיתוח גישות חדשות לבחינת הקשר בין חלבונים ופרקטלים וזאת מתוך מטרה כפולה: (א') ניסיון להמשיך ולבסס את נקודת המבט הפרקטלית על חלבונים באמצעות קבלת תוצאות מוכרות בדרכים חדשות; (ב') ניסיון לחשוף רבדים נוספים ועמוקים יותר העומדים בבסיס הגישה הקיימת. במהרה הסתמן כי גם כאן, כמו במקרים

רבים אחרים, ניתן יהיה להפיק תועלת רבה משימוש בגישות אשר אקראיות עומדת בבסיסן וזאת לשם הבנת תהליכים שלכאורה אין בינם ובין אקראיות ולא כלום. מתברר כי ניתן ללמוד רבות על אלמנטים מבנים ודינמיים המאפיינים רשת באמצעות מעקב אחר חלקיק המבצע הילוך אקראי על אותה הרשת. ניתן לייצג מולקולת חלבון כרשת של חומצות אמינו המחוברות זו לזו. אם נאפשר לחלקיק לבצע הילוך אקראי על רשת זו תשתקפנה התכונות הפרקטליות, במידה והן אכן קיימות, בסטטיסטיקה של המסלולים האקראיים אותם נדגום.

במחקר שבו ניתחנו את התכונות הסטטיסטיות של הילוך אקראי המבוצע על גבי מולקולת חלבון ישמנו רעיון זה הלכה למעשה [49]. מהמחקר עלו המסקנות הבאות: (א') כמצופה מהילוך אקראי על מבנה פרקטלי, הילוך אקראי על מולקולת חלבון הוא הילוך אקראי אנומלי. התכונות הפרקטליות של המבנה משתקפות בתוצאות הנצפות ומאששות תצפיות קודמות; (ב') קיימים קשרים רבים בין הילוך אקראי על רשת והדינמיקה הויברציונית של הרשת עצמה. מקשרים אלו עולה שהילוך אקראי אנומלי שקול לדינמיקה ויברציונית אנומלית. מכיוון והתכונות הסטטיסטיות של ההילוך האקראי מושפעות אך ורק ממבנה החלבון עליו מתבצע ההילוך, אנומליות רבות בדינמיקה הויברציונית של חלבונים הן למעשה תוצאה ישירה של מבנה פרקטלי; (ג') ניתן לחזות אספקטים מסוימים בהתנהגות של מולקולת חלבון בודדת בהסתמך על האנלוגיה בין מהלכים אקראיים ודינמיקה ויברציונית. אם לא דיי באלו, עם התקדמות המחקר גילינו שניתן, באמצעות כלים שפותחו במקור לצורך חקר הדינמיקה של חלבונים, ללמוד גם על בעיות יסודיות בתורה הסטוכסטית של המהלכים האקראיים.

בעיית זמן הגעה ראשון היא בעיה הצצה באופן טבעי במגוון רחב של יישומים כגון: ראקציות כימיות, התפשטות מחלות, תהליכי חיפוש ועוד... בכל אותם מקרים נשאלת, בגרסאות שונות, השאלה הבאה: "כמה זמן לוקח למהלך אקראי להגיע מנקודה א' לנקודה ב' ?". שאלה זו היא שאלה הסתברותית מטבעה כיוון וזמן ההגעה הראשון הוא זמן אקראי. לאחרונה, נתגלה קשר בין זמן ההגעה הממוצע של מהלך אקראי על רשת והדינמיקה הויברציונית של אותה הרשת כאשר היא מצומדת לאמבט חום. פריצת הדרך נעוצה בכך שנמצאה בעיה מכנית (פיזיקלית) השקולה לבעיה ההסתברותית המקורית. אנו הסתמכו על הקשר הנ"ל כאשר הדגמנו כיצד ניתן, מתוך שיקולים פיזיקליים בסיסיים הנוגעים לתכונות המכניות של רשתות, לפתור את בעיית זמן ההגעה הראשון הממוצע על רשת פרקטלית [83].

הממוצע הוא נציגה הבולט ביותר של ההתפלגות. יחד עם זאת, פעמים רבות אנו מתעניינים דווקא בסטיות מן הממוצע ולעתים אף בשאלות הקשורות להתפלגות כולה. מסיבה זו פעלנו על מנת להרחיב בצורה משמעותית ולהעמיד על רגליים מתמטיות איתנות את האנלוגיה הקיימת בין בעיית זמן ההגעה הראשון והדינמיקה של רשתות אלסטיות המצומדות לאמבט חום [84]. ההתפלגות המלאה של זמן ההגעה הראשון וכן הסתברויות אכלוס ומעבר, הובעו כולן במונחים של פונקציות קורלציה ויברציוניות ובין היתר התברר שקיימת שקילות מלאה בין בעיית החזרה לראשית ועקרון אי היציבות של לנדאו ופיירלס. רשתות פרקטליות נלמדו כמקרה בוחן והוכח שהממד הספקטרלי הוא הקובע האם זמן ההגעה הממוצע מייצג היטב את ההתפלגות ממנה נלקח.

מלבד חלבונים, פרקטלים משמשים מודל למערכות רבות אחרות, כאשר בין היתר ניתן למנות: זכוכיות בטמפרטורה נמוכה, חומרים פורוזיביים, אגרגטים של קולואידים ואף דנ"א. ניסיונות פיזור, בהם נמדדות קורלציות בזמן ובמרחב, מאפשרים את אפיונם של חומרים אלו. בקונטקסט זה, נודעה חשיבות רבה לגודל המכונה "גורם המבנה". ואולם, בעוד שגורם המבנה הסטטי מובן היטב, גורם המבנה הדינמי מובן הרבה פחות. בפרט, לא קיימת תיאוריה החוזה את הדעיכה בזמן של גודל זה בהסתמך על עקרונות ראשוניים. אנו חישבנו את גורם המבנה הדינמי עבור רשתות פרקטליות ומצאנו כי הוא מתאפיין בדעיכה אנומלית [68]. יתר על כן, הפונקציה המתמטית המאפיינת את הדעיכה היא מסוג אקספוננט מתוח והאקספוננט עצמו נקבע ע"י הממד הפרקטלי והממד הספקטרלי המאפיינים את המבנה. התאוריה אותה פיתחנו מאפשרת קביעה של הממדים האופייניים בהסתמך על מדידה ניסיונית של גורם

המבנה הדינמי.

התפתחויות טכנולוגיות מהשנים האחרונות אפשרו את ביצועם של ניסיונות פורצי דרך, בהם נמדד גורם המבנה הדינמי של תמיסה דלילה המכילה את החלבונים המוגלובין ומיוגלובין. התוצאות הצביעו על דעיכה אנומלית של גורם המבנה הדינמי כאשר בחלון הזמן הרלוונטי נצפתה דעיכה מסוג אקספוננט מתוח. אנו חקרנו את גורם המבנה הדינמי של החלבונים הנזכרים לעיל ושל חלבונים נוספים [72]. המחקר, שנערך תוך אינטגרציה בין תוצאות אנליטיות, סימולציות מחשב ותצפיות ניסיוניות, אישש את הסברה שהמבנה הפרקטלי הוא הסיבה לדעיכה האנומלית של גורם המבנה. מאידך התברר, שההתאמה בין התאוריה והניסיון אינה מושלמת, בין היתר כיוון שחלבונים אינם מערכות פרקטליות אידאליות והמודל הפרקטלי נכון עבורם רק בקירוב. בדומה לכל אידאליזציה, גם על נקודת המבט המוצגת בחיבור זה חלות מגבלות וסייגים. יחד עם זאת, רוצה אני להאמין שהממצאים שתוארו לעיל מדגישים את התועלת הרבה הנובעת מאימוץ הגישה המתוארת בין דפים אלו ואת חשיבותה ככלי המשלים את הגישות הקיימות בתחום.

4 Proteins and Fractals

4.1 Proteins

This section will provide a short and informal introduction to proteins. The main aim is to provide the reader with a general idea about proteins and the structure and function of proteins in particular.

The word protein comes from the Greek word "prota", meaning "of primary importance". These molecules were first described and named by the Swedish chemist Jöns Jakob Berzelius in 1838. However, proteins' central role in living organisms was not fully appreciated until 1926, when James B. Sumner showed that the enzyme urease was a protein[1]. Proteins are of prime importance in organisms and participate in numerous processes within cells. Many proteins are enzymes that catalyze biochemical reactions, and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscles, and the proteins in the cytoskeleton, which forms a system of scaffolding that maintains the cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle.

4.1.1 Biochemistry

Proteins are large organic compounds, made of amino acids [2]. In proteins, amino acids are first arranged in a linear chain, joined together by peptide bonds that are formed between the carboxyl atom of one amino acid and the amine nitrogen of another. Once linked in a chain, each amino acid is called a residue and the linked series of carbon, nitrogen, and oxygen atoms are known as the protein backbone. Due to the chemical structure of the individual amino acids, the protein chain has directionality. The end of the protein with a free carboxyl group is known as the C-terminus or carboxyl terminus, while the end with a free amino group is known as the N-terminus or amino terminus.

4.1.2 Biosynthesis

The sequence of amino acids in a given protein is dictated by a corresponding gene, encoded in the genetic code. Genes are actually segments of DNA that are first transcribed into pre-messenger RNA and then translated into proteins. This idea is part of the central dogma of molecular biology [2]. The unique amino acid sequence is specified by the nucleotide sequence of the gene encoding the protein. The genetic code is a set of three-nucleotide sets called codons and each three-nucleotide combination stands for an amino acid. For instance, CAG stands for Glutamine and CCC stands for Proline. Because DNA contains four nucleotides, the total number of possible codons is 64. Hence, there is some redundancy in the genetic code and some amino acids are specified by more than one codon. The amino acid sequence of a protein is also referred to as the primary structure of the protein. The first protein to be sequenced was insulin, by Frederick Sanger, who won the Nobel Prize for this achievement in 1958.

4.1.3 Structure and Function

Proteins are known to naturally fold into three dimensional structures which are favored under physiological conditions and are hence termed 'native structures'. The native protein structure usually has functional relevance, a protein that has not folded into its native structure may not function properly. Biochemists often refer to four distinct aspects of a protein structure:

- Primary structure: the amino acid sequence.
- Secondary structure: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha helix and beta sheet. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.
- Tertiary structure: the overall shape of a single protein molecule; the spatial relationship of secondary structures to one another. Tertiary structure is generally stabilized by non-local interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulphide bonds, and even post-translational modifications.
- Quaternary structure: the shape or structure that results from the interaction of more than one protein molecule, usually called protein subunits in this context, which function as part of the larger assembly or protein complex.

Proteins are flexible molecules. In addition to these levels of structure, proteins may shift between several related structures while they perform their biological function. In the context of these functional rearrangements, tertiary or quaternary structures are usually referred to as 'conformations', and transitions between them are called 'conformational changes'. Such changes are often induced by the binding of a substrate molecule to an enzyme's active site which is the physical region of the protein that participates in chemical catalysis. In solution all proteins also undergo variations in structure due to collisions with other molecules. These variations are usually referred to as 'thermal vibrations'.

The loss of native conformation at extreme pH values, at high temperatures and in the presence of denaturing substances (such as urea) is known as 'denaturation'. The fact that a denatured protein can spontaneously return to its native conformation was demonstrated for the first time with ribonuclease, a digestive enzyme. Discovering the native structure of a protein can provide important clues about how a protein performs its function. Common experimental methods of structure determination include X-ray crystallography and NMR spectroscopy, both of which can produce information at the atomic resolution. The first protein structures to be solved included hemoglobin and myoglobin, by Max Perutz and Sir John Cowdery Kendrew, respectively, in 1958 [3, 4]. The structures were determined via X-ray diffraction analysis and granted their discoverers the 1962 Nobel Prize in Chemistry. Today there are more than 75,000 structures in the Protein Data Bank (PDB) [5].

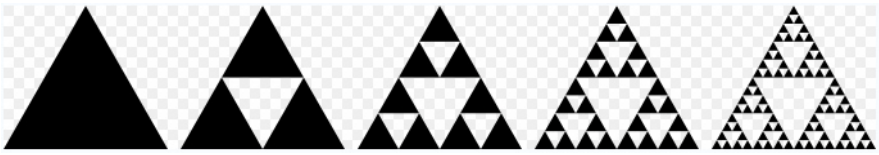


Figure 2: One can construct the Sierpinski gasket recursively as follows. Start with any triangle in a plane, the canonical Sierpinski gasket uses an equilateral triangle with a base parallel to the horizontal axis (left image). Shrink the triangle by half, make two copies, and position the three shrunken triangles such that each triangle touches the two other triangles at a corner (second image from the left). Repeat the second step with each of the smaller triangles *ad infinitum* ...

4.2 Fractals

This section will provide a short and informal introduction to fractals. The main aim is to provide the reader with a general idea about fractals and fractal dimensions. A well known fractal, the Sierpinski gasket, will serve as an example. We start we a common definition for a fractal:

- Fractal - "A rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced size copy of the whole" Benoît Mandelbrot [6].

The term fractal was coined by Benoît Mandelbrot in 1975 and was derived from the Latin word fractus, meaning 'broken' or 'fractured'. A fractal as a geometric object generally has the following features [7]:

1. Fine structure at arbitrarily small scales.
2. Self similarity, be it exact or statistical.

The above features may be rather puzzling encountered for the first time. We will now try to shed some light on their meaning through a simple example.

4.2.1 The Sierpinski Gasket

The Sierpinski gasket, also called the Sierpinski triangle, is a fractal, named after Waclaw Sierpinski who described it in 1915. The Sierpinski gasket is one of the basic examples of self-similar sets. Figure (2) illustrates the recursive construction of the Sierpinski gasket. Following the basic step described in the caption of figure (2) an infinite number of times, it is clear why the Sierpinski gasket exhibits fine structure at arbitrarily small scales. In addition, the Sierpinski gasket is clearly self similar since it is constructed from three smaller Sierpinski gaskets that are exact miniature versions of the original gasket.

Figure 3: Starting with the small Sierpinski gasket on the left and treating it as a basic unit, N denotes the number of basic units required to construct each of the gaskets above. L denotes the scale factor between the side of the basic unit and the side of the gaskets above. Noting that in the $n-th$ stage $N = 3^n$ and $L = 2^n$, one may verify that $N = L^{ln(3)/ln(2)}$ by taking the natural logarithm of both sides.

4.2.2 The Fractal Dimension

Fractals as we shall exemplify shortly, push the intuitive notion of dimension beyond its naive use. Consider for example an equilateral triangle (which we often think of as two dimensional) with side a. The area of such a triangle is $S = \frac{\sqrt{3}a^2}{4}$, the length of its perimeter is $L = 3a$. The properties of non zero area and finite perimeter length appear very natural. Indeed, many two dimensional geometrical objects such as the disk and square exhibit them. One the other hand, the Sierpinski gasket does not.

Recall the construction of the Sierpinski gasket and assume we start with an equilateral triangle of side a. The area of the geometrical object obtained after the $n-th$ recursive stage is $S_n = \frac{\sqrt{3}a^2}{4} \cdot \left[\frac{3}{4}\right]^n$. The area of the Sierpinski gasket is the limit of S_n as n tends to infinity, we hence conclude that this area is zero! The perimeter length of the geometrical object obtained after the $n-th$ recursive stage is $L_n = 3a \cdot \left[\frac{3}{2}\right]^n$. The perimeter length of the Sierpinski gasket is the limit of L_n as n tends to infinity, we hence conclude that this length is infinite! What is the dimensionality of the Sierpinski gasket then? Is it a two dimensional object with zero area or a bounded one dimensional object with infinite length? The answer to this question depends on the definition of the ambiguous term: 'dimension'.

Suppose that a fractal consists of N identical parts that are similar to the entire fractal with a scale factor of L, is there a relation between N and L? Figure (3) shows that for the Sierpinski gasket such a relation exists, indeed for this fractal object: $N = L^{ln(3)/ln(2)}$. We define the fractal dimension d_f to be the exponent of L in the power law relation between N and L. From this definition it follows that the fractal dimension of the Sierpinski gasket is

$d_f = ln(3)/ln(2) \simeq 1.585$. Repeating the procedure described in figure (3) with a regular equilateral triangle (instead of the Sierpinski gasket) we would come to the conclusion that in the $n-th$ stage $N = 4^n$ and $L = 2^n$. In this case it is clear that $N = L^2$ and $d_f = 2$. Indeed, the fractal dimension is a generalization of the 'regular' dimension we are all accustomed with and hence for regular two dimensional objects the fractal dimension coincides with the 'regular' dimension.

The definition of the fractal dimension as presented in the paragraph above is somewhat limiting. We may generalize this definition for a general object composed out of generic 'building blocks' as follows: (i) Locate the object's center of mass and draw a sphere of radius r around it; (ii) Denote the number of 'building blocks' enclosed by this sphere $n(r)$; (iii) If $n(r)$ scales as the radius to some power, define that power to be the fractal dimension. For regular objects, the fractal dimension calculated in this way coincides with the our 'regular' notion of dimension. The fractal dimension of the Sierpinski gasket is independent of the way we calculate, it hence remains $ln(3)/ln(2)$.

4.3 The Mass Fractal Dimension of Proteins

The procedure described above is useful in computing the mass fractal dimension of proteins as well. The mass fractal dimension serves as a measure of the density in which a protein fills space. Describing the way in which the mass fractal dimension is calculated is most convenient using a three dimensional example. Draw a sphere of radius r whose center coincides with the protein's center of mass. Calculate the mass $M(r)$ of all amino acids enclosed by this sphere, increase r and calculate again. Do this several times and if $M(r)$ scales as r^{d_f} the exponent d_f is called the fractal dimension. Enright *et al.* preformed a large scale analysis of the mass fractal dimension in proteins, calculating the fractal dimension for 200 proteins [8]. It has been found that proteins can be described as mass fractals whose mass fractal dimension d_f is close to 2.5 (with a statistical standard deviation of about 0.2). In light of the fact that for a regular three dimensional lattice, $d_f = 3$, the smaller fractal dimension ($2 < d_f < 3$) found for proteins is a result of a sparser fill of the space. And so, although they are embedded in a three dimensional space, proteins are in a sense entities of lower and fractional dimensionality.

5 Modeling Protein Dynamics

The amino acid sequence, or primary structure, of a protein is its unique fingerprint. While it is believed that structure is encoded in the amino acid sequence the former can not always be reliably predicted based on the latter. Recent advances in structural biology have highlighted the tremendous importance of protein structure in determining function and allowing for proper functionality [9]. Protein structure has therefore emerged as an important source of additional information required for understanding the molecular basis of observed biological activities. Proteins however, are flexible molecules that fluctuate continuously due to collisions with the molecules of the solvent and other surrounding molecules. The conformational dynamics of protein molecules is encoded in their structures and is often a critical element of their function. Structural knowledge alone is hence insufficient. In order to fully decipher the connection between structure and function, vibrational dynamics must be understood as well.

5.1 All Atom Models and Molecular Dynamics Simulations

A major endeavor in recent years has been to develop models and methods for simulating the dynamics of proteins, and relating the observed behavior to experimental data. Computer based molecular dynamics simulations provide links between structure and dynamics by enabling the exploration of molecular motions and trajectories. In the past two decades, the size of system addressable with computer simulations has gradually increased. Enhanced computer power has now reached memory and speed requirements sufficient to treat explicitly solvated proteins; new and more efficient techniques to sample configurational space have been proposed [10, 11]; and force-field parameterization, optimized through a continuous and ongoing validation process, has become accurate enough to explain and predict experimental results [12, 13].

The first molecular dynamics simulation of a protein was reported in 1977 and consisted of a 9.2-ps trajectory of a small protein in vacuum [14]. Today, ten and even hundred nanosecond simulations of proteins in an explicit aqueous environment are feasible [15] and yet, many questions of interest are still beyond the reach of molecular dynamics. Some processes, such as rearrangements upon ligand binding, protein-protein docking and protein folding, occur on timescales which are much longer than few nanoseconds. Other processes involve large protein structures or supramolecular assemblies for which a full atom treatment is beyond memory and/or time limitations. In addition, force fields used in molecular dynamics are often aimed at attaining a high degree of realism which inevitably comes at the expense of mathematical tractability. There is therefore no hope of building a tractable analytical model that possesses a similar degree of realism and while this should not interfere with ongoing computer based simulations it may very well impair our ability to obtain a deep and fundamental understanding of the observed results.

All of the above join the fact that atomic resolution is not always within the reach of the experimental method at hand (for example consider Cryo-electron microscopy). In these cases, computer simulations at the atomic level are somewhat of an overkill and a simplified description in which many microscopic degrees of freedom are grouped together, or averaged over, is much more natural. The 'integration' of a large number of degrees of freedom into a smaller number that represent the overall dynamics at a somewhat coarser resolution, is sometimes referred to as 'coarse graining'. Coarse-grained approaches are ubiquitous in physics and have enjoyed much interest in the context of proteins dynamics as well.

5.2 Coarse-Grained Elastic Network Models

Most analytical treatments of protein dynamics entail a compromise between physical realism and mathematical tractability. The challenge is to identify a simple, yet physically plausible, model that retains properties of interest and experimental relevance [16]. While accurate sampling of the conformational space is a challenging problem for macromolecular systems, the study of protein dynamics benefits from a great simplification. Under physiological conditions, each protein is characterized by a unique native state structure and is functional only when folded in this particular way. While the motions of macromolecules in solution are quite complex and involve transitions between an astronomical number of conformations, those of proteins near native state conditions are much simpler, as they are confined to a subset of conformations near the folded state. These conformations usually share the same overall fold, secondary structural elements, and even tertiary contacts within individual domains. A typical example is the closed and open forms of enzymes which are adopted upon ligand binding and unbinding respectively.

Exploring the fluctuation dynamics of proteins near native state conditions is a first step towards the understanding of the molecular basis and mechanisms of their function. Elastic network models exploit the existence of a well defined native state in order to present a simplified description of protein dynamics. In these models a protein is represented as a network of beads (each amino acid can be represented by a single bead or more) connected by elastic springs, as shown in figure (4). Thus, elastic network models replace detailed atomic potentials with simple harmonic potentials between interacting beads. While the exact nature of the interaction between beads may vary between different models, elastic network models are usually linear models amenable to normal mode analysis [17, 18, 19].

In comparison to their relative simplicity, elastic network models were shown to perform surprisingly well in predicting large scale collective motions of proteins. Concerted motions of large groups of atoms, residues, or even entire domains were shown to arise repeatedly and underlie the biological function of many different proteins. It was noted that structural changes are often dominated by a small set of distinct deformations that require small energy ascents in the multidimensional energy landscape. These deformations are in fact low

Figure (4): Elastic network models provide a bead-spring representation of proteins.

frequency (soft) normal modes. Normal mode analysis can hence be used in order to identify them and predict preferable cooperative motions in proteins.

One possible explanation regarding the success of elastic network models is the apparent robustness of large-scale collective motions. Provided that the topology of the bead spring network remains the same, collective motions were shown to be relatively insensitive to the details of the model used [20, 21, 22, 23, 24, 25, 26, 27]. This surprising result has motivated researchers to study in detail the interplay between purely topological constraints, defined by the 3D structure, and the collective dynamics of proteins. A paradigmatic model used for that purpose is the Gaussian Network Model.

5.3 The Gaussian Network Model

Following Tirion's original work [20], the Gaussian Network Model (GNM) proposed by Bahar *et al.* [21], utilizes the harmonic approximation and is widely applied because it yields results in agreement with X-ray spectroscopy experiments. The GNM considers proteins to be elastic networks whose beads correspond to the positions of the alpha-carbons in the native structure and the interactions among beads are modeled as identical harmonic springs. An interaction between two nodes exists if the beads are separated by a distance less than R_c, a parameter known as the interaction cutoff. The cutoff distance is usually taken in the range $6\mathring{A} - 7\mathring{A}$, based on the radius of the first coordination shell around residues observed in PDB structures [28, 29]. The only information required to implement the method is knowledge of the native structure.

The GNM is defined by the quadratic Hamiltonian equation

$$H_{GNM}=\sum_i \frac{(\overrightarrow{P_i})^2}{2M} + \frac{\gamma}{2}\sum_{i,j>i}\Lambda_{ij}(\triangle\overrightarrow{R_i} - \triangle\overrightarrow{R_j})^2. \quad (1)$$

The first term represents the kinetic energy of the system, γ is the spring force constant which is assumed to be homogeneous, $\overrightarrow{R_i}$ and $\triangle\overrightarrow{R_i}$ are the instantaneous position and the displacement with respect to equilibrium position, $\overrightarrow{R_i^0}$, of the i-th C_α atom respectively. Λ is the network connectivity matrix with the following entries: $\Lambda_{ij} = 1$ if $i \neq j$ and the distance $\left|\overrightarrow{R_i^0} - \overrightarrow{R_j^0}\right|$ between two C_α atoms, in the native conformation, is below the cutoff R_c. In all other cases$\Lambda_{ij} = 0$. Physically, this means that in addition to changes in inter-residue distances, any change in the direction of the inter-residue vector is also being resisted or penalized in the GNM potential. As is clear from the separable Hamiltonian, fluctuations in the GNM are isotropic.

5.3.1 Classical Mechanics of the GNM

Consider a general elastic network of masses and harmonic springs. It is well known that one of the characteristics of such a network is a set of normal modes and a corresponding set of eigenfrequencies. A normal mode of an oscillating system is a pattern of motion in which all parts of the system move sinusoidally with the same frequency. The frequencies of the normal modes of a system are known as its natural frequencies or eigenfrequencies.

In order to find the normal modes and eigenfrequencies of an elastic network within the framework of the GNM we use Lagrangian mechanics (Hamiltonian mechanics will of course yield the same results). The classical Lagrangian of the GNM reads

$$\mathcal{L} = \sum_i \frac{M(\dot{\overrightarrow{R_i}})^2}{2} - \frac{\gamma}{2}\sum_{i,j>i}\Lambda_{ij}(\triangle\overrightarrow{R_i} - \triangle\overrightarrow{R_j})^2 \ . \quad (2)$$

Written in a less compact form

$$\begin{cases} \mathcal{L} = \sum_i \frac{M\triangle\dot{X_i}^2}{2} + \frac{M\triangle\dot{Y_i}^2}{2} + \frac{M\triangle\dot{Z_i}^2}{2} \\ -\frac{\gamma}{2}\sum_{i,j>i}\Lambda_{ij}\left[(\triangle X_i - \triangle X_j)^2 + (\triangle Y_i - \triangle Y_j)^2 + (\triangle Z_i - \triangle Z_j)^2\right] \ , \end{cases} \quad (3)$$

it is clear that this Lagrangian is three times degenerate and separable into the variables: $\triangle X_i$,$\triangle Y_i$,$\triangle Z_i$. It is hence sufficient to concentrate on

$$\mathcal{L} = \sum_i \frac{M\triangle\dot{X_i}^2}{2} - \frac{\gamma}{2}\sum_{i,j>i}\Lambda_{ij}(\triangle X_i - \triangle X_j)^2 \ . \quad (4)$$

To continue further we first note that

$$\sum_{i,j>i}\Lambda_{ij}(\triangle X_i - \triangle X_j)^2 = \sum_{i,j>i}\Lambda_{ij}(\triangle X_i^2 - 2\triangle X_i\triangle X_j + \triangle X_j^2) = \sum_{i,j}\Gamma_{ij}\triangle X_i\triangle X_j \ , \quad (5)$$

where the matrix Γ is defined as follows

$$\Gamma_{ij} = \begin{cases} -1 & if\ i \neq j\ and\ R^0_{ij} \leq R_c \\ 0 & if\ i \neq j\ and\ R^0_{ij} > R_c \\ \sum\limits_k \Lambda_{ik} & if\ i = j \end{cases} . \tag{6}$$

We can hence write the Lagrangian in the following form

$$\mathcal{L} = \sum_i \frac{M \triangle \dot{X_i}^2}{2} - \frac{\gamma}{2} \sum_{i,j} \Gamma_{ij} \triangle X_i \triangle X_j \ . \tag{7}$$

Recalling the Euler Lagrange equations of motion ($\frac{\partial \mathcal{L}}{\partial \triangle X_i} = \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \triangle \dot{X}_i}$) we get an equation for each and every node in the network

$$M \triangle \ddot{X}_i = -\gamma \sum_j \Gamma_{ij} \triangle X_j \ . \tag{8}$$

This set of N equations can be written in matrix form

$$M \triangle \ddot{\vec{X}} = -\gamma \Gamma \triangle \vec{X} \ . \tag{9}$$

Substituting an oscillatory solution, $\triangle \vec{X} = \vec{a} e^{i\omega t}$, we get an eigenvalue equation for the matrix Γ

$$\Gamma \vec{a} = \frac{M\omega^2}{\gamma} \vec{a} \ . \tag{10}$$

We conclude that the eigenfrequencies of the this elastic network are, up to a proportionality factor, the square root of the eigenvalues of the matrix Γ.

5.4 The Spectral Dimension

5.4.1 The Spectral Dimension of Fractals

The spectral dimension d_s governs the density of low frequency normal modes on a fractal. More precisely, denoting the density of modes with frequency ω, $g(\omega)$, the scaling relation $g(\omega) \sim \omega^{d_s - 1}$ holds for low frequencies. A related quantity is the cumulative density of states: $G(\omega) = \int_0^\omega g(\omega') d\omega'$. $G(\omega)$ counts the number of modes with frequency less than ω and the scaling relation $G(\omega) \sim \omega^{d_s}$ holds for low frequencies. Numerically $G(\omega)$ is a monotonically increasing step function and it is obtained directly, without further approximations, given the set of eigenfrequencies that characterize the system. It is hence convenient utilize the relation $G(\omega) \sim \omega^{d_s}$ when trying to estimate the spectral dimension.

The spectral dimension d_s is hence a quantity directly related to the vibrational dynamics of a fractal. It is well known that for an infinite regular two dimensional lattice the spectral dimension coincide with the 'regular' dimension,

i.e. $d_s = 2$. In fact, this is true for regular lattices in any euclidean dimension d, the density of modes obeys the Debye relation $g(\omega) \sim \omega^{d-1}$ for low frequencies and $d_s = d$ [30]. The spectral dimension is hence another generalization of the 'regular' dimension we all seem to know.

In similarity to the fractal dimension, the spectral dimension of the Sierpinski gasket is also smaller than 2 and yet the two dimensions are distinct. The spectral dimension is given by $2ln(3)/ln(5) \simeq 1.365$ and is hence smaller than the fractal dimension [31, 32]. Figure (5) describes two different elastic networks of masses and springs, the first is a finite Seirpinski gasket and the second is a regular square lattice. Figure (6) describes the spectral analysis of these networks for a finite Sierpinski gasket of 6561 beads and a 81×81 regular square lattice.

5.4.2 The Spectral Dimension of Proteins

The GNM models a protein as an elastic network. Burioni *et al.* [33] computed the spectral dimension, for a set of 57 proteins, within the GNM framework. This can be done by building and diagonalizing the matrix Γ[1] defined in Eq. 6 for each protein of interest. By doing so, Burioni *et al.* obtained the set of vibrational eigenfrequencies $\{\omega_1, \omega_2, ..., \omega_N\}$, plotting $ln(G(\omega))$ vs. $ln(\omega)$, it was found that low frequency region of the spectrum clearly exhibits a power-law behavior. Moreover, for most proteins it was found that $d_s < 2$, although higher values of the interaction cutoff R_c lead to higher values of d_s [33]. This result is surprising since proteins are usually considered regular three dimensional objects for which the spectral dimension coincides with the dimension of the embedding space, i.e., $d_s = 3$. And yet, it was found that the spectral dimension of proteins is considerably lower than that of the embedding space, in similarity to what was found for the mass fractal dimension and in accord with their fractal-like nature. Once again, proteins were found to be entities of lower and fractional dimensionality.

[1] The eigenfrequencies of the the elastic network are, up to a scale factor, the square root of the eigenvalues of the matrix Γ. One should not be bothered with the proportionality factor since we are only interested in the scaling law of $G(\omega)$ with ω.

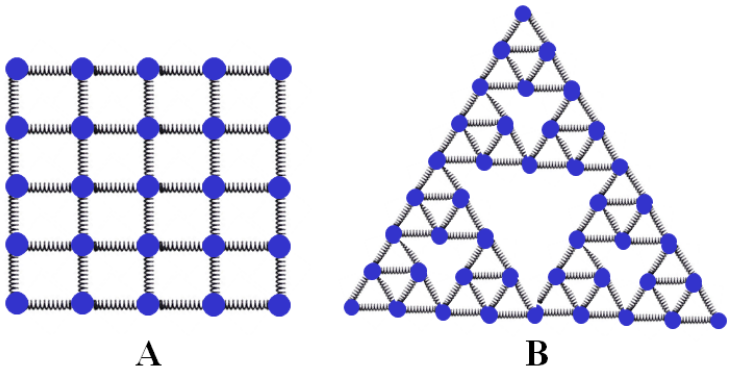


Figure 5: Two different elastic networks of masses and springs. Left - a regular square lattice as an elastic network, every node is connected with springs to its nearest neighbors. Right - the Sierpinski gasket as an elastic network, the beads here are the vertices of the original gasket at each recursive step.

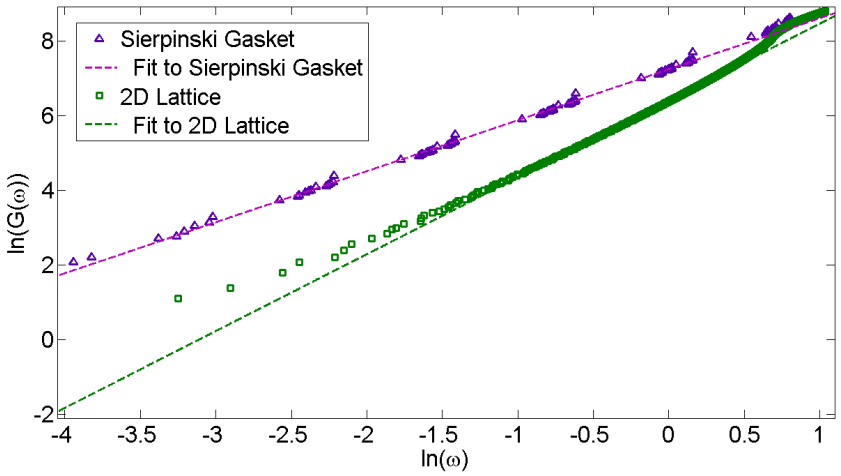


Figure 6: Spectral analysis of two finite elastic networks. For each elastic network, we found the set of vibrational eigenfrequencies $\{\omega_1, \omega_2, ..., \omega_N\}$ that characterize it and then plotted $ln(G(\omega))$ vs. $ln(\omega)$. Low frequency regions of the cumulative density of modes $G(\omega)$ clearly exhibit a power law behavior. Dashed lines indicate best fits to these regions, the slopes correspond to the spectral dimension and it is clearly visible that $d_s^{square} > d_s^{gasket}$. Numerical values for the spectral dimensions are found to be $d_s^{square} = 2.05$ and $d_s^{gasket} = 1.36$. These values agree with the theoretical values for the corresponding infinite networks.

6 Thermal Vibrations - Static Properties

Proteins vibrate due to collisions with the molecules of the solvent and with other surrounding molecules. In the common jargon of statistical physics it is said that a protein is in contact with a 'thermal bath'. It is now time to consider the effect of this coupling on the vibrational dynamics of proteins. We start by considering a basic quantity: the static vibrational mean square displacement.

6.1 The Static Vibrational Mean Square Displacement

One may wonder to what extent does the GNM provide a reliable description of the vibrational dynamics of proteins? The mean square displacement of a residue from its equilibrium position is experimentally measurable. In X-ray crystallography this quantity is directly related to measurable B-factors and in NMR experiments this quantity is simply the root mean-square difference between different NMR models. A thermodynamic analysis of the GNM [17, 21] provides us with a formula for the theoretical value of the experimentally measurable B-factors mentioned above

$$B_i \equiv \frac{8\pi^2}{3} < \triangle \vec{R}_i^2 >= \frac{8\pi^2 k_B T}{\gamma} \left[\Gamma^{-1}\right]_{ii} \ , \tag{11}$$

where the subscript i stands for the $i-th$ residue and the pointy brackets denote the thermal average. It is thus possible to compare between theoretical and experimental results.

Starting with the paper that introduced the GNM [21], several studies have demonstrated that the fluctuations predicted by the GNM are in good agreement with experimental B-factors. It is interesting to note that in a recent study [34] conducted on a set of 64 nonhomologous proteins, each containing a structure solved by NMR and X-ray crystallography. The GNM predictions for mean square fluctuation yielded a correlation of 0.59 with X-ray data and a distinctively better correlation (0.75) with NMR data. The higher correlation between GNM and NMR data, compared to that between GNM and X-ray B factors, was shown to arise from the differences in the spectrum of modes accessible in solution and in the crystal environment. Namely, large amplitude motions sampled in solution are restricted, if not inaccessible, in the crystalline environment.

6.2 Landau-Peierls Instability

A classical result obtained by Peierls in 1934 [35] provides a thermodynamic explanation for the instability of low dimensional crystalline structures. The argument, based on harmonic vibrational dynamics, shows that the mean square displacement of any structural unit at finite temperature diverges in the thermodynamic limit when the lattice dimension is 1 or 2. Indeed, when such a quantity exceeds the order of magnitude of the lattice spacing, the structure behaves as

a liquid and the crystalline order makes no longer sense. Interestingly, the result of Peierls holds even if anharmonic terms are present since the instability is present at any finite temperature and in particular in the low-temperature regime in which the contribution from anharmonic terms is negligible. On the other hand, on real finite structures far from the thermodynamic limit, the crystalline order is stable if the mean-square displacement does not exceed the lattice spacing. The maximum stability size at room temperature is so small for $d = 1$ that it make no sense speaking of 1-dimensional crystals, while for $d = 2$ the logarithmic divergence is slow enough to allow the existence of mesoscopic two dimensional crystals.

6.2.1 Landau-Peierls Instability of Fractals and Proteins

The classical result of Peierls can be extended for non crystalline structures such as fractals and proteins. It turns out that stability in the thermodynamic limit is possible if and only if $d_s > 2$ [33]. As mentioned previously, proteins are characterized by a spectral dimension which is usually smaller than two and a stability problem is hence inevitable [33]. A possible explanation regarding why this is so, together with the solution of the stability problem will be discussed in chapter 9. In what follows we will revisit the vibrational mean square displacement and try to understand the connection between topology and instability.

The average mean square displacement of a network bead in the GNM is given by

$$< \triangle \vec{R}^2 >= \frac{\sum_{i=1}^{N} < \triangle X_i^2 > + < \triangle Y_i^2 > + < \triangle Z_i^2 >}{N} = \frac{3\sum_{i=1}^{N} < \triangle X_i^2 >}{N} \ , \quad (12)$$

where we have averaged over all the beads in the network and used the fact that the GNM is three times degenerate. For a given structure, it is possible to calculate the thermal averages that appear in Eq. 12 by use of Eq. 6. This method requires however an ad-hoc calculation of the matrix Γ^{-1}. It is therefore advantageous to consider an alternative path that sheds a more general light on this problem.

6.2.2 Normal Coordinates

As was described in 5.3.1, in the GNM the dynamics of an elastic network is fully described by the following set of equations

$$M \triangle \ddot{X}_i = -\gamma \sum_j \Gamma_{ij} \triangle X_j \ , \quad (13)$$

which in matrix form read

$$M \triangle \ddot{\vec{X}} = -\gamma \Gamma \triangle \vec{X} \ . \quad (14)$$

A brief look at these equations reveals that they are coupled. The dynamics of the i-th node, described by the deviation from equilibrium $\triangle X_i$, depends not

only on $\triangle X_i$ itself but also on other beads $\{\triangle X_j\}$. Although the above description is very natural, since the coordinates used are the actual deviations from equilibrium, it leads to a rather complicated set of equations. It is sometimes beneficial to describe the system differently using a special set of coordinates called normal coordinates. Describing the system with normal coordinates leads to N uncoupled equations of motion and great mathematical simplicity.

The matrix Γ defined in 5.3.1 is real and symmetric[2] by definition. One of the basic theorems concerning such matrices is the finite-dimensional spectral theorem, which says that any symmetric matrix whose entries are real can be diagonalized by an orthogonal matrix[3]. More explicitly: to every symmetric real matrix Γ there exists a real orthogonal matrix A such that $D = A^{-1}\Gamma A \equiv A^T\Gamma A$ is a diagonal matrix. Every symmetric matrix is thus, up to choice of an orthonormal basis, a diagonal matrix. Another way of stating the real spectral theorem is that the eigenvectors of a symmetric matrix are orthogonal. More precisely, a matrix is symmetric if and only if it has an orthonormal basis of eigenvectors.

Letting A be the real orthogonal matrix that diagonalizes the matrix Γ. We define a new set of coordinates $\{\triangle U_i\}$ using the old set of coordinates $\{\triangle X_i\}$ by the orthogonal transformation

$$\triangle\vec{X} = A\triangle\vec{U} \ . \tag{15}$$

We are now able to obtain the equations of motion for the new coordinates $\{\triangle U_i\}$, noting that $A\triangle\ddot{\vec{U}} = \triangle\ddot{\vec{X}} = -\frac{\gamma}{M}\Gamma\triangle\vec{X} = -\frac{\gamma}{M}\Gamma A\triangle\vec{U}$ we get

$$\triangle\ddot{\vec{U}} = -\frac{\gamma}{M}A^{-1}\Gamma A\triangle\vec{U} = -\frac{\gamma}{M}D\triangle\vec{U} \ , \tag{16}$$

where D is a diagonal matrix whose entries are the eigenvalues of the matrix Γ, i.e the set $\left\{\frac{M}{\gamma}\omega_i^2\right\}$. As promised we got a set of N uncoupled equations which read

$$\triangle\ddot{U}_i = -\omega_i^2\triangle U_i \ . \tag{17}$$

The normal coordinate $\triangle U_i$ obeys the equation of motion for a simple harmonic oscillator with angular frequency ω_i.

6.2.3 Deriving the Instability Criterion

Normal coordinates allow us to calculate the static vibrational mean square displacement with relative ease. We first note that

$$\sum_i < \triangle X_i^2 >= \sum_i < \left[\sum_j A_{ij}\triangle U_j\right]^2 >=< \sum_i\sum_j\sum_k A_{ij}A_{ik}\triangle U_j\triangle U_k > \tag{18}$$

[2]In linear algebra, a symmetric matrix is a square matrix Γ, that is equal to its transpose $\Gamma = \Gamma^T$. The entries of a symmetric matrix are symmetric with respect to the main diagonal (top left to bottom right), so if the entries are written as a_{ij}, then $a_{ij} = a_{ji}$.

[3]In matrix theory, a real orthogonal matrix is a square matrix A whose transpose is its inverse: $A^TA = AA^T = I$. A real square matrix is orthogonal if and only if its columns form an orthonormal basis of the Euclidean space R_n with the ordinary Euclidean dot product, which is the case if and only if its rows form an orthonormal basis of R_n.

$$=<\sum_j\sum_k\sum_i A_{ij}A_{ik}\triangle U_j\triangle U_k>=<\sum_j\sum_k\delta_{jk}\triangle U_j\triangle U_k>=\sum_j<\triangle U_j^2>\ ,$$

where we have used the fact that the columns of the matrix A are orthonormal. Calculating the thermal average $<\triangle U_i^2>$ now boils down to calculating the mean square displacement of a simple harmonic oscillator coupled to a thermal bath. After factoring out the integration over momenta, $<\triangle U_i^2>$ is given by

$$<\triangle U_i^2>=\frac{\int_{-\infty}^{\infty}\triangle U_i^2 e^{-\frac{M\omega_i^2\triangle U_i^2}{2k_BT}}d\triangle U_i}{\int_{-\infty}^{\infty}e^{-\frac{M\omega_i^2\triangle U_i^2}{2k_BT}}d\triangle U_i}=\frac{k_BT}{M\omega_i^2}=\frac{k_BT}{\gamma l_i}\ , \tag{19}$$

where l_i denotes the i-th eigenvalue of the matrix Γ and γ is the GNM spring constant. Denoting the density of eigenvalues of the matrix Γ, $g(l)$, we replace summation with integration and obtain

$$<\triangle\vec{R}^2>=\frac{3}{N}\sum_{i=1}^{N}<\triangle U_i^2>=\frac{3k_BT}{N\gamma}\int_{l_{min}>0}^{l_{max}}\frac{g(l)}{l}dl\ , \tag{20}$$

where l_{min} denotes the smallest positive eigenvalue and l_{max} the largest.

In order to evaluate $<\triangle\vec{R}^2>$ in the thermodynamic limit i.e., when $N\to\infty$, we must first say something about the scaling properties of $g(l)$ and l_{min} . First, we note that $g(l)$ is an extensive quantity and hence scales with N, this is readily understood from the normalization condition $\int_{l_{min}>0}^{l_{max}}g(l)dl=N$. Second, as was mentioned in 5.4.1 the spectral dimension d_s governs the density of low frequency normal modes on a fractal/protein. More precisely, denoting the density of modes with frequency ω : $g(\omega)$, the scaling relation $g(\omega)\sim\omega^{d_s-1}$ holds for low frequencies. Since $l\backsim\omega^2$ we deduce the following scaling law, $g(l)\sim l^{\frac{d_s-2}{2}}$, which holds for low eigenvalues. Third, the lowest positive eigenfrequency ω_{min} corresponds to the lowest wave number k_{min}. The lowest wave number possible is limited by the size[4] of the elastic network $k_{min}\sim\frac{1}{R_g}\sim\frac{1}{N^{1/d_f}}$, using the known dispersion relation for fractals[31] $\omega\sim k^{\frac{d_f}{d_s}}$, we conclude that $l_{min}\sim\omega_{min}^2\sim k_{min}^{\frac{2d_f}{d_s}}\sim N^{-\frac{2}{d_s}}$.

The above scaling relations allow us to compute the integral in Eq. 20 and conclude that

$$<\triangle\vec{R}^2>\sim\frac{k_BT}{\gamma}N^{\frac{2}{d_s}-1}\ . \tag{21}$$

[4]The radius of gyration [36], $R_g=\sqrt{\sum_i m_i r_i^2/\sum_i m_i}$, is a characteristic length scale. Here the sum runs over all network beads, m_i is the mass of the $i-th$ node and r_i the distance from the center of mass.

Examining Eq. 21 it is clear that when taking the limit $N \to \infty$, the average mean square displacement diverges if and only if $d_s \leq 2$.[5] We have thus proved that stability in the thermodynamic limit is possible if and only if $d_s > 2$. This results extends the Peierls theorem form crystalline structures characterized by an integer spectral dimension to fractal structures characterized by a broken spectral dimension.

[5] For $d_s = 2$ one must recalculate the integral and find that the divergence is actually logarithmic.

7 Thermal Vibrations - Dynamic Properties

Having calculated the static vibrational mean square displacement we now turn our attention to dynamical properties of elastic networks in contact with a thermal bath. The approach taken here is based on the Langevin equation which we review for the sake of completeness. The equation is then applied in tandem with the GNM in order to provide a basic theoretical framework for the study of the vibrational dynamics of proteins.

7.1 The Langevin Equation

In statistical physics, a Langevin equation is a stochastic differential equation describing Brownian motion in a potential. For a single particle in solution Newton's second law reads

$$m\frac{d^2x}{dt^2} = -\zeta\frac{dx}{dt} - \frac{\partial U}{\partial x} + h(t) \ . \tag{22}$$

Here the first term on the right hand side is the friction force which is assumed to take a standard form of being opposite in direction and proportional to the velocity. The second term is the force exerted as a consequence of the external potential and the third term is a random force that represents the sum of the forces due to collisions with surrounding particles.

Let us now rewrite Eq. 22 in the following way

$$\frac{m}{\zeta}\frac{d^2x}{dt^2} + \frac{dx}{dt} = -\frac{1}{\zeta}\frac{\partial U}{\partial x} + \frac{g'(t)}{\zeta} \equiv -\frac{1}{\zeta}\frac{\partial U}{\partial x} + g(t) \ , \tag{23}$$

where we have defined $g(t) \equiv \frac{h(t)}{\zeta}$. Our next step is an approximation, treating very small and light weight particles we will drop the inertial term $\frac{m}{\zeta}\frac{d^2x}{dt^2}$ assuming it is negligible and obtain

$$\frac{dx}{dt} = -\frac{1}{\zeta}\frac{\partial U}{\partial x} + g(t) \ . \tag{24}$$

We will refer to Eq. 24 as the Langevin equation. Equation 24 describes the motion of a single Brownian particle in the overdamped limit, solving it one can (in principle) obtain the trajectory of such a particle.

The approximation we have performed can be further justified by considering a particle immersed in some solvent and moving under the influence of a constant external force $F = -\frac{\partial U}{\partial x}$. Let us denote the velocity

$$v(t) = \frac{dx(t)}{dt} \ , \tag{25}$$

the equation of motion for $v(t)$ is given by

$$\frac{m}{\zeta}\frac{dv(t)}{dt} + v(t) = \frac{F}{\zeta} + g(t) \ . \tag{26}$$

For simplicity let us factor out the random force by taking an ensemble average of both sides of the equation thus obtaining an equation for the average velocity

$$\frac{d < v(t) >}{dt} + \frac{\zeta}{m} < v(t) >= \frac{F}{m} \ . \tag{27}$$

Multiplying both sides by $e^{\frac{t\zeta}{m}}$ and integrating from zero to t we are able to solve for $< v >$ and obtain

$$< v(t) >= e^{\frac{-t\zeta}{m}} \int_0^t \frac{F}{m} e^{\frac{t'\zeta}{m}} dt' = \frac{F}{\zeta} \left[1 - e^{\frac{-t\zeta}{m}} \right] \ , \tag{28}$$

where we have assumed that the particle was at rest at time zero. We see that the velocity approaches an asymptotic value of $\frac{F}{\zeta}$ exponentially fast and that the characteristic relaxation time is $\tau = \frac{m}{\zeta}$. Dropping the inertial term in the first place we would have simply gotten

$$< v >= \frac{F}{\zeta} \ , \tag{29}$$

which can be interpreted physically as an immediate response to the force. It is now clear that if the relaxation time $\tau = \frac{m}{\zeta}$ is small, dropping the inertial term is a good approximation! In the case of small particles (atoms,molecules,colloidal particles, etc...) immersed in liquid, the relaxation time τ is indeed very small supporting the validity of the approximation we have made.

7.1.1 From Trajectories to Probability Distributions

Examining many trajectories one can generate the probability distribution for the position of a particle at time t. For example starting the particle from a given origin and following its trajectory up to some time t, one can record the position $x(t)$. Repeating this process many times will yield many different values for $x(t)$. By use of a histogram, one can generate an empirical probability distribution for the position at time t. It is possible to show [37] that if the probability distribution of $g(t)$ is Gaussian and characterized by

$$\begin{cases} < g(t) >= 0 \\ < g(t)g(t') >= \frac{2k_B T}{\zeta} \delta(t - t') \ , \end{cases} \tag{30}$$

then $\Psi(x,t)$ (the probability distribution of $x(t)$) determined by the Langevin equation satisfies the Smoluchowski equation

$$\frac{\partial \Psi(x,t)}{\partial t} = \frac{\partial}{\partial x} \frac{1}{\zeta} \left[k_B T \frac{\partial \Psi(x,t)}{\partial x} + \Psi(x,t) \frac{\partial U}{\partial x} \right] \ . \tag{31}$$

In other words, if $g(t)$ is a Gaussian random variable with zero mean and variance $\frac{2k_B T}{\zeta}$ and if $g(t)$ and $g(t')$ are independent for $t \neq t'$ then the above statement holds.

As a concrete example consider the Brownian motion of a free particle (no external potential) for which the Langevin equation reads

$$\frac{dx}{dt} = g(t) \ . \tag{32}$$

If the particle is at x_0 at time $t = 0$, its position at time t is given by

$$x(t) = x_0 + \int_0^t g(t')dt' \ . \tag{33}$$

From the above we deduce that $x(t) - x_0$ is a linear combination independent Gaussian random variables. Recalling that the sum of independent Gaussian random variables is a Gaussian random variable itself, the probability distribution of $x(t)$ may be written as

$$\Psi(x,t) = \frac{1}{\sqrt{2\pi B(t)}} exp\left[-\frac{(x - A(t))^2}{2B(t)}\right] \ , \tag{34}$$

where

$$\begin{cases} A(t) =< x(t) > \\ B(t) =< (x(t) - A(t))^2 > \end{cases} \ . \tag{35}$$

The mean is readily calculated from

$$A(t) =< x(t) >= x_0 + \int_0^t < g(t') > dt' = x_0 + \int_0^t 0 \cdot dt' = x_0 \ , \tag{36}$$

and for the variance we have

$$B(t) =< \left(\int_0^t g(t')dt'\right)\left(\int_0^t g(t'')dt''\right) >= \int_0^t \int_0^t < g(t')g(t'') > dt'dt'' \tag{37}$$

and it therefore follows that

$$B(t) = \int_0^t \int_0^t \frac{2k_BT}{\zeta}\delta(t - t')dt'dt'' = \frac{2k_BT}{\zeta}t = 2Dt \ . \tag{38}$$

We thus conclude that

$$\Psi(x,t) = \frac{1}{\sqrt{4\pi Dt}} exp\left[-\frac{(x - x_0)^2}{4Dt}\right] \tag{39}$$

which is exactly (check by direct differentiation) the solution for the Smoluchowski equation

$$\frac{\partial \Psi(x,t)}{\partial t} = D\frac{\partial^2 \Psi(x,t)}{\partial x^2} \ . \tag{40}$$

An important conclusion is that the mean square displacement (from the origin) of a Brownian particle is given by $2Dt$ and is hence linear with time.

7.2 Langevin Equation for the Harmonic Oscillator

Consider a Brownian particle moving under the following potential

$$U(x) = \frac{1}{2}kx^2 \ . \tag{41}$$

The equation of motion for this particle is given by

$$\frac{dx(t')}{dt'} = -\frac{k}{\zeta}x(t') + g(t') \ . \tag{42}$$

In order to get a formal solution for $x(t)$ we multiply both sides by $e^{\frac{k}{\zeta}t'}$ and do some algebra to get

$$\frac{dx(t')}{dt'}e^{\frac{k}{\zeta}t'} + \frac{k}{\zeta}x(t')e^{\frac{k}{\zeta}t'} = \frac{d}{dt'}\left[x(t')e^{\frac{k}{\zeta}t'}\right] = g(t')e^{\frac{k}{\zeta}t'} \ . \tag{43}$$

Integrating from $-\infty$ to t we get

$$\left[x(t')e^{\frac{k}{\zeta}t'}\right]_{t'=-\infty}^{t'=t} = \int_{-\infty}^{t} g(t')e^{\frac{k}{\zeta}t'}dt' \tag{44}$$

and after applying the following boundary condition[6]

$$x(t' = -\infty) \cdot e^{-\frac{k}{\zeta}\cdot\infty} = 0 \ , \tag{45}$$

we conclude that

$$x(t) = \int_{-\infty}^{t} g(t')e^{\frac{k}{\zeta}(t'-t)}dt' \ . \tag{46}$$

In order to calculate moments, we note that for the mean position we have

$$A(t) = < x(t) > = \int_{-\infty}^{t} < g(t') > e^{\frac{k}{\zeta}(t'-t)}dt' = 0 \ . \tag{47}$$

We will now aim at finding an expression for the mean square displacement from the origin $< (x(t) - x(0))^2 >$ and obtain the variance of $x(t)$ as a by product. We start with the time correlation function of $x(t)$

$$< x(t)x(0) > = \int_{-\infty}^{t} dt_1 \int_{-\infty}^{0} dt_2 e^{\frac{k}{\zeta}(t_1+t_2-t)} < g(t_1)g(t_2) > \ , \tag{48}$$

[6] It is also possible to solve under the initial condition $x(t = 0) = x_0$, in that case $\left[x(t')e^{\frac{k}{\zeta}t'}\right]_{t'=0}^{t'=t} = \int_0^t g(t')e^{\frac{k}{\zeta}t'}dt'$ and we have $x(t) = x_0e^{-\frac{k}{\zeta}t} + \int_0^t g(t')e^{\frac{k}{\zeta}(t'-t)}dt'$.

for which, by use of Eq. 46, we get

$$< x(t)x(0) >= \int_{-\infty}^{0} e^{\frac{k}{\zeta}(2t_1-t)} \frac{2k_BT}{\zeta} dt_1 = \left[e^{\frac{k}{\zeta}(2t_1-t)} \frac{k_BT}{k} \right]_{t_1=-\infty}^{t_1=0} = \frac{k_BT}{k} e^{-\frac{k}{\zeta}t} \ . \tag{49}$$

Here we assumed that $t > 0$ and used the fact that $< g(t_1 > 0)g(t_2) >\equiv 0$ since $t_2 < 0$. Similarly if $t < 0$ we get

$$< x(t)x(0) >= \int_{-\infty}^{t} e^{\frac{k}{\zeta}(2t_1-t)} \frac{2k_BT}{\zeta} dt_1 = \left[e^{\frac{k}{\zeta}(2t_1-t)} \frac{k_BT}{k} \right]_{t_1=-\infty}^{t_1=t} = \frac{k_BT}{k} e^{\frac{k}{\zeta}t} \ , \tag{50}$$

and we may hence conclude that

$$< x(t)x(0) >= \frac{k_BT}{k} e^{-\frac{k}{\zeta}|t|} \ . \tag{51}$$

Letting $t = 0$ we have

$$< x(0)x(0) >=< x^2 >= \frac{k_BT}{k} \tag{52}$$

which coincides with the known result obtained from statistical mechanics by use of the Boltzmann distribution $\psi_{eq} \propto exp(-kx^2/2k_BT)$.

Interestingly, the right hand site of Eq. 52 is also the variance of $x(t)$. Indeed

$$B(t) =< (x(t) - A(t))^2 >=< x(t)^2 >= \int_{-\infty}^{t}\int_{-\infty}^{t} < g(t')g(t'') > e^{\frac{k}{\zeta}(t'+t''-2t)} dt'dt'' \tag{53}$$

and hence

$$B(t) = \int_{-\infty}^{t} \frac{2k_BT}{\zeta} e^{\frac{k}{\zeta}(2t'-2t)} dt' = \frac{k_BT}{k} \ . \tag{54}$$

The mean square displacement $< (x(t) - x(0))^2 >$ can now be easily computed by noting that

$$< (x(t)-x(0))^2 >=< x(t)^2 > + < x(0)^2 > -2 < x(t)x(0) >= 2\left[< x(t)^2 > - < x(t)x(0) > \right] \tag{55}$$

and it follows immediately that

$$< (x(t) - x(0))^2 >= \frac{2k_BT}{k} \left[1 - e^{-\frac{k}{\zeta}|t|} \right] \ . \tag{56}$$

Here, unlike the case of free diffusion, for long times the mean square displacement is bounded by $\frac{2k_BT}{k}$. The bound is approached exponentially fast

with a characteristic relaxation time $\tau = \frac{\zeta}{k}$. Considering the opposite limit $|t| \to 0$ (very short times) we have (to first order)

$$< (x(t) - x(0))^2 >= \frac{2k_BT}{k} \left[1 - 1 + \frac{k}{\zeta} |t| \right] = 2D |t| \ . \tag{57}$$

Indeed in this limit the particle has yet to "feel" the harmonic potential and we expect regular diffusion. Since $x(t)$ is a linear sum of Gaussian random variables and hence Gaussian itself. As we have already found the mean and variance, we can also write an expression for the probability distribution of $x(t)$:

$$\Psi(x,t) = \frac{1}{\sqrt{2\pi B}} exp \left[-\frac{(x - A(t))^2}{2B(t)} \right] = \frac{1}{\sqrt{\frac{2\pi k_BT}{k}}} exp \left[-\frac{kx^2}{2k_BT} \right] \tag{58}$$

which is exactly the Boltzmann distribution. As we have given the particle an infinite amount of time to equilibrate with the potential well, we could have guessed this result beforehand.

7.3 Langevin Equation for the Gaussian Network Model

Coupling an elastic network to a thermal bath the Langevin equation, within the framework of GNM, reads

$$\frac{d\triangle\vec{X}}{dt} = -\frac{\gamma}{\zeta}\Gamma\triangle\vec{X} + \vec{g}(t) \ , \tag{59}$$

where the random forces are characterized by

$$\begin{cases} < \vec{g}(t) >= \vec{0} \\ < g_i(t)g_j(t') >= \frac{2k_BT}{\zeta}\delta_{ij}\delta(t-t') \quad \forall i,j \in \{1...N\} \ . \end{cases} \tag{60}$$

Rewriting this equation in terms of the normal coordinates defined in Eq. 15 we obtain

$$\frac{d\triangle\vec{U}}{dt} = -\frac{\gamma}{\zeta}D\triangle\vec{U} + \tilde{g}(t) \ , \tag{61}$$

where $D = A^{-1}\Gamma A$ is a diagonal matrix whose entries are the eigenvalues of the matrix Γ, i.e the set $\{\omega_1^2, \omega_2^2, ..., \omega_N^2\}$ and $\tilde{g}(t) \equiv A^{-1}\vec{g}(t)$. The random forces are now characterized by:

$$\begin{cases} < \tilde{g}(t) >=< A^{-1}\vec{g}(t) >= A^{-1} < \vec{g}(t) >= \vec{0} \\ < \tilde{g}_i(t)\tilde{g}_j(t') >= \frac{2k_BT}{\zeta}\delta_{ij}\delta(t-t') \quad \forall i,j \in \{1...N\} \ , \end{cases} \tag{62}$$

where we have noted that: $< \tilde{g}(t)\tilde{g}(t')^T >=< A^{-1}\vec{g}(t)\vec{g}(t')^T A >= A^{-1} < \vec{g}(t)\vec{g}(t')^T > A = \frac{2k_BT}{\zeta}\delta(t-t')A^{-1}IA = \frac{2k_BT}{\zeta}\delta(t-t')I$. The equation of motion for $\triangle U_i$ now reads

$$\frac{d\triangle U_i}{dt} = -\frac{\gamma\omega_i^2}{\zeta}\triangle U_i + \tilde{g}_i(t) \ , \tag{63}$$

and it is thus uncoupled from $\triangle U_j$ for $i \neq j$. Moreover, $\triangle U_i$ obeys the Langevin equation of a simple harmonic oscillator with $k = \gamma\omega_i^2$. The results we have obtained in 7.2 are thus of prime importance in this case as well.

8 Mapping Random Walks and Thermal Vibrations

Mapping two different physical problems onto one another has been proven very useful in physics. Examples are the mapping between the Schrodinger and diffusion equations [38], the mapping between lattice-gas and Ising models [39], the mapping between quantum field theories and critical phenomena [39] and the mapping between random walks and electric networks [40]. Here we present a brief introduction to the well established mapping between random walks and thermal vibrations [41].

8.1 Markov Chains

In mathematics, a Markov chain, named after Andrey Markov, is a stochastic process with the Markov property [42]. At each point in time the state of a system may change. If a sequence of states has the Markov property, then every future state is independent of every past states given the present state. In other words, having the Markov property means the next state depends solely on the present state and not on previous states.

A bit more formally, a Markov chain is a sequence of random variables $X_1, X_2, X_3, ...$ with the Markov property. Given the present state, the future and past states are independent and we we write

$$P(X_{n+1} = x_{n+1}|X_n = x_n, ..., X_1 = x_1) = P(X_{n+1} = x_{n+1}|X_n = x_n) \ . \quad (64)$$

The possible values of X_i come form a countable set S called the state space of the chain and $P(X_{n+1} = x_{n+1}|X_n = x_n)$ denotes the probability that $X_{n+1} = x_{n+1}$ given that $X_n = x_n$. Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states. Time-homogeneous Markov chains (or, Markov chains with time-homogeneous transition probabilities) are processes where

$$P(X_{n+1} = x|X_n = y) = P(X_n = x|X_{n-1} = y) \ , \quad (65)$$

for all $n \in N$ and all $x, y \in S$.

As a concrete example of a simple, time-homogeneous, Markov chain consider the following scenario. The probabilities of weather conditions given the weather on the preceding day are represented by a transition matrix

$$P = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \quad (66)$$

The matrix P represents a weather model in which a sunny day is 90% likely to be followed by another sunny day, and a rainy day is 50% likely to be followed by another rainy day. The columns can be labeled 'sunny' and 'rainy' respectively, and the rows can be labeled in the same order. P_{ji} is the probability that, if a given day is of type i, it will be followed by a day of type j. Notice that the columns of P sum to 1 as should be.

8.2 Random Walks and the Gaussian Network Model

Markov chains can be used to describe random walks on graphs. Replacing beads with nodes and springs with edges the Gaussian network, as in the the GNM, turns into a three dimensional graph. Relating transition probabilities with each edge completes the metamorphosis and turns the Gaussian network into a Markov chain. The Markov chain created in this way has a number of states that is equal to the number nodes in the network (amino acids along the protein backbone).

A simple way to create such a chain is to think of each spring in the GNM as an edge and attribute transition probabilities as follows

$$\begin{cases} P_{ji} = \frac{1}{d_i} & R_{ij}^0 \leq R_c \\ P_{ji} = 0 & otherwise \end{cases} \tag{67}$$

Here P_{ji} denotes the probability to transit from state i to state j , d_i denotes the coordination number of the $i-th$ node, R_{ij}^0 the distance between residue i and j and R_c the cutoff distance of the GNM. Note that as should be: $\sum_j P_{ji} = 1$ since by definition there are only d_i non-zero terms which all equal $\frac{1}{d_i}$. The conditional probability matrix $P = \{P_{ji}\}$, also called the Markov transition matrix, defines a Markov chain over the network.

Suppose the probability of initiating the Markov propagation process at node i is $p_i(0)$. Then, the probability of reaching node j in one step is $P_{ji} \cdot p_i(0)$. In matrix notation, the probability of ending up on any of the residues after one step is given by the distribution $\vec{p}(1) = P \cdot \vec{p}(0)$. Iterating this process we conclude that after k steps

$$\vec{p}(k) = P^k \cdot \vec{p}(0) \tag{68}$$

where $\vec{p}(k) = [p_1(k), p_2(k), .., p_N(k)]$ represents the n-dimensional vector of the probabilities to reside at node $1 \leq j \leq N$ after k steps.

8.3 Random Walks and Thermal Vibrations

In continuous time the change in occupation probabilities follows the master equation

$$\frac{d\vec{p}(t)}{dt} = W\vec{p}(t) \ , \tag{69}$$

with the following solution [42]

$$\vec{p}(t) = exp[tW]\vec{p}(0) \ , \tag{70}$$

which replaces Eq. 68. Here W is the transition rates matrix whose matrix element W_{ij} is the rate of transition from state j to state i . Note that in order to obey probability conservation we must have $\sum_{i \neq j} W_{ij} = -W_{jj}$, i.e the rate at which probability exits the state j must be equal to the total rate at which probability enters the different states $i \neq j$ given it emanated at j. We note

that this picture is equivalent to a picture in which the random time it takes a random walker to leave site j is exponentially distributed with rate $-W_{jj}$ and where immediately after this time the random walker has probability $\frac{W_{ij}}{-W_{jj}}$ to be found at state $i \neq j$ [42].

A natural way to define a continuous time random walk on a Gaussian network is to choose for W the GNM matrix $-\frac{\gamma}{\zeta}\Gamma$ defined previously, by doing so we get

$$\frac{d\vec{p}(t)}{dt} = -\frac{\gamma}{\zeta}\Gamma\vec{p}(t) \ . \tag{71}$$

Equation 71 share resemblance with Eq. 59 that describes the thermal vibrations of an elastic network coupled to a thermal bath. Indeed, taking the thermal average of Eq. 59 the equation for $< \triangle\vec{X} >$ reads

$$\frac{d < \triangle\vec{X} >}{dt} = -\frac{\gamma}{\zeta}\Gamma < \triangle\vec{X} > \ . \tag{72}$$

In other words, the time evolution of the probability vector $\vec{p}(t)$ identifies with the time evolution of the average displacements vector $< \triangle\vec{X} >$. This important relation will be advanced upon later on.

Moreover, given an initial condition for the occupancy probabilities at time zero, they can be expressed it in terms of the eigenvectors of the matrix Γ

$$\vec{p}(0) = \sum_{j=1}^{N} a_j \vec{A}_j \ , \tag{73}$$

where the coefficient a_i is found by multiplying both sides of the equation by the row vector $(\vec{A}_i)^T$ and the use of orthonormality $((\vec{A}_i)^T \cdot \vec{A}_j = \delta_{ij})$

$$a_i = (\vec{A}_i)^T \cdot \vec{p}(0) \ \ \forall i. \tag{74}$$

The solution at time $t > 0$ can now be written in following form

$$\vec{p}(t) = e^{-\frac{\gamma t}{\zeta}\Gamma}\vec{p}(0) = \sum_{j=1}^{N} a_j e^{-\frac{\gamma\omega_j^2 t}{\zeta}} \vec{A}_j \ , \tag{75}$$

and it is now clear that the set of vibrational eigenfrequencies $\{\omega_1, \omega_2, ..., \omega_N\}$ (and consequently the vibrational density of states) plays an important role here as well.

9 The Articles

In this chapter we will review the articles and course of study. The chapter is divided to three parts: *foundations, applications and ramifications.* A major concern we have had when we started our study was regarding the extent to which proteins may be considered fractals. We have thus dedicated a special effort in order to solidify the *foundations* of the fractal approach, explore new aspects of it and sharpen the boundaries and limitations of the fractal-protein analogy. From here we have turned to study possible implications and *applications* of the fractal-like nature of proteins. As an interesting by product, we discovered our work had unexpected *ramifications.*

9.1 Foundations

Thinking about proteins such as enzymes and antibodies, uniqueness and specificity immediately come to mind. However, as scientists we strive for general simple rules that describe the objects we study, for example, the ideal gas equation:

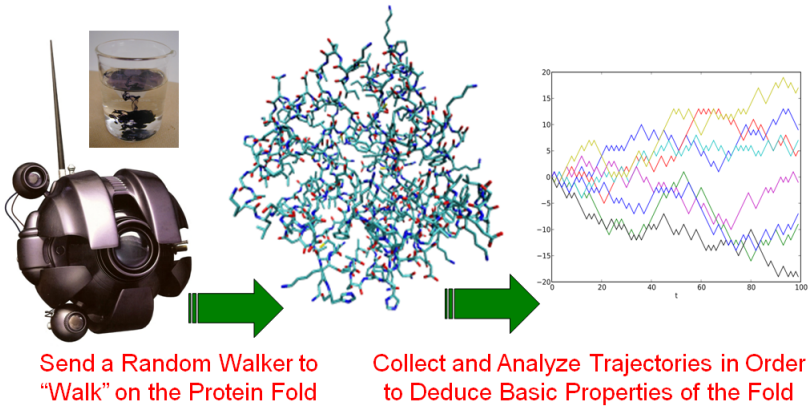
$$PV = Nk_BT \ . \tag{76}$$

In this spirit and despite the large diversity, we took the challenge of searching for a non-trivial common denominator unifying the entire proteomic fauna. Adopting a fractal perspective on proteins, essential information regarding protein topology and dynamics can be treated within a single unifying framework. As with fractals, proteins can also be characterized by a small set of broken dimensions. The fractal dimension d_f characterizes the way in which a protein fills space [8] and the spectral dimension d_s characterizes the way in which a protein vibrates [33].

Two seemingly conflicting properties of native proteins are known to coexist. While proteins must keep their specific native structure thermally stable, the native fold displays the ability to perform flexible motions that allow for proper function [43, 44, 45, 46]. Revisiting the Landau-Peierls instability criterion we have demonstrated how the delicate balance between protein stability and flexibility imposes a constraint on the relation between the number of amino acids along the protein backbone N, the spectral dimension, d_s, and the fractal dimension d_f [47]. This constraint leads to a surprising conclusion: proteins are fractals of a very special kind; Statistically, they all obey an 'equation of state' in the fractal parameter space.

Combining physical modeling together with bioinformatical tools, we have analyzed protein structure data, acquired via X-Ray crystallography experiments, and validated the 'equation of state', mentioned above, for a data set of 543 proteins taken from the protein data bank [47]. Later on, we have demonstrating universality by extending our analysis and confirming the validity of our observations for a data set which was ten-fold larger [48].

The way in which the fractal and spectral dimensions are defined (see section 4.3 and section 5.4 respectively) is quite common and is based on the space



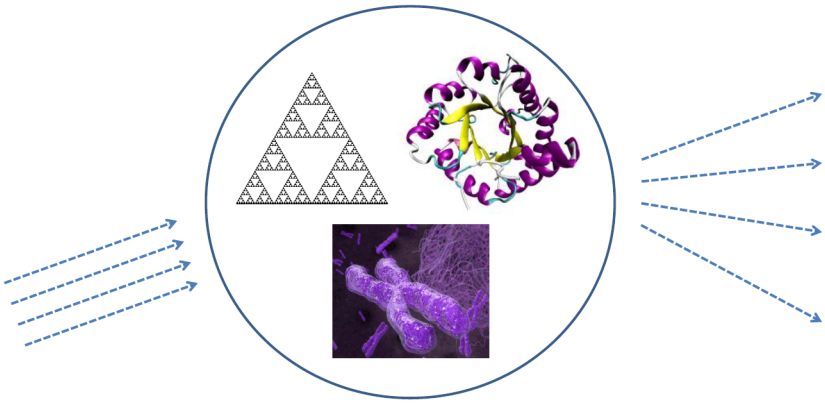
filling and mechanical properties of proteins. While the number of proteins for which a fractal and spectral dimensions were calculated increased from few hundreds (prior to our studies) to several thousands, we sought for further establishment of the fractal approach towards proteins. Studies have shown that it is possible to deduce the innate properties of networks by analyzing the stochastic trajectories of a random walker on them. For fractal networks, it is possible to characterize basic quantities such as: the probability to be at the origin and the mean square displacement (MSD) from the origin, by use of the fractal and spectral dimensions. Structure, vibrational dynamics and random walks are hence strongly connected. Recalling that proteins may be represented as networks of masses and springs, known relations between random walks and vibrational dynamics bridge the gap between network science and protein dynamics at equilibrium.

We have used random walks (or equivalently diffusion) in order to probe protein structure and dynamics. Conducting a large scale study of diffusion on more than 500 proteins we find it to be anomalous, in similarity to many other fractals. Our findings provide an alternative and independent confirmation of the fractal-like nature of proteins [49].

9.2 Applications

The fractal perspective on protein structure and dynamics is useful in providing a unified approach towards a number of seemingly unrelated phenomena. Proteins exhibit strange and anomalous dynamics displaying non-Debye density of vibrational states [33, 47, 48, 50, 51], anomalous spread of vibrational energy [52, 53], large conformational changes [54, 55, 56] and non-exponential decay of time correlation functions [57]. The anomalous behavior may, in principle, stem from various factors affecting the energy landscape under which proteins vibrate and yet, among all these, a place of honor is saved for the structure-dynamics interplay.

Taking advantage of known [41] and newly derived [49, 58] relations between



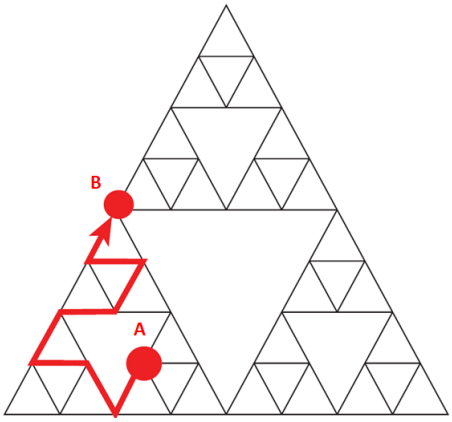
Scattering experiments are a classical approach to the characterization of matter. We present a theory that explains the decay of the dynamic structure factor of fractal-like object such as proteins and chromatin.

vibrational dynamics and diffusion, we have demonstrated the equivalence between observed anomalies in diffusion on proteins and anomalies in their vibrational dynamics. Diffusion is a method sensitive to the structural and topological features of proteins and them alone. Anomalies in the vibrational dynamics of proteins are hence a direct result of their fractal-like structure [49].

The duality between diffusion and vibrational dynamics allowed us to make, single molecule level, experimentally testable predictions [49]: (i) The time dependent vibrational mean square displacement of an amino acid was predicted to be subdiffusive; (ii) The thermal variance in the instantaneous distance between amino acids was predicted to scale as a power law of the equilibrium distance between them. In addition, mean first passage time analysis was offered as a practical tool that may aid in the identification of amino acid pairs involved in large conformational changes.

Proteins are not the only naturally occurring fractal-like structures. Naturally occurring fractals are ubiquitous [6] and fractal models have been used to describe the dynamics of low temperature glasses and porous materials [59], sol-gel branched polymer clusters [60, 61, 62], colloidal aggregates [63] and DNA [64, 65, 66]. Scattering experiments, in which one is able to simultaneously probe correlations in space and time, allow the characterization of fractal structures. A key player in these experiments is the structure factor [37, 67]. While the static structure factor of fractals is well understood [37, 59, 67], dynamic structure factor calculations are much more limited.

Motivated by the lack of an adequate theory, we have studied the dynamic structure factor of large vibrating fractal networks [68]. Extending a well known, 'text book', result from the spatial regime to the spatial-temporal regime, we have shown that the temporal decay of the dynamic structure factor is dominated by the spatially averaged mean square displacement of a network bead.



The first passage time problem on the Sierpinski gasket. How long does it take a random walker to get from point A to point B?

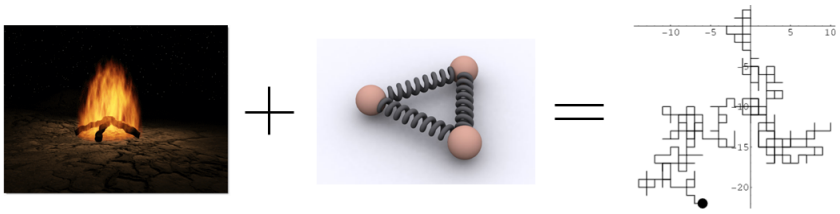
The latter evolves subdiffusively in time and as a result, the decay of the dynamic structure factor follows a stretched exponential law.

Recent advancements in high-resolution inelastic neutron scattering, available with neutron spin-echo spectroscopy, have turned this approach useful in the analysis of biomolecule flexibility and vibrational dynamics [69, 70]. Neutron spin-echo studies that measure the dynamic structure factor have been recently performed on horse heart myoglobin and bovine hemoglobin in solutions [71] and the temporal decay of the dynamic structure factor was well fitted by a stretched exponential. This observation is in accord with our theory and has further motivated us to analyze the dynamic structure factor of several specific proteins as a case study [72].

9.3 Ramifications

In the context of our study, the mapping between vibrational dynamics and random walks was utilized, first and foremost, as a practical tool. It was only deep into the study where we have come to realize that a recently noted analogy, between the mean first passage time (MFPT) problem and the thermal variance in the distance between two amino acids [58], opens new lines of theoretical investigation.

The first passage time problem plays an important role in situations such as transport in disordered media [73, 74], neuron firing [75], spread of diseases [76] and target search processes [77, 78, 79, 80]. In the MFPT problem one is concerned with the following question: "on average, how long does it take a random walker to travel from point A to point B?". The problem is, of course, probabilistic in nature and its general solution in the case of self similar networks was derived a short while ago using probabilistic methods [81, 82]. Interestingly, the recently noted analogy between the MFPT and thermal vibrations provides a non-probabilistic approach towards the solution of the MFPT problem [83]. It



Probabilistic questions concerning random walks can be answered via a non probabilistic approach - analysis of thermal vibrations.

turns out that on fractal networks, the solution to this problem is readily obtained using 'a vibrational shortcut'. One first casts the original (probabilistic) problem onto an analogous (and simple) vibrational problem. After solving the vibrational problem, the solution is mapped back and results are interpreted probabilistically.

The MFPT is the most distinct representative of the first passage time problem and yet, one may wonder regarding vibrational analogs of other related quantities as well. As mentioned previously, Chennubhotla *et al.* expressed the MFPT in terms of thermal vibrational correlation functions [58]. However, the line of argumentation used by them could not be carried on to higher moments of the first passage time, nor to its full distribution or any additional quantities. We have revisited the mapping between thermal vibrations and random walks. Advancing upon it, we examined general networks and were able to express random walk occupation probabilities, first passage time distributions and passage probabilities between nodes in terms of thermal vibrational correlation functions [84]. In the special case of fractal networks, the spectral dimension was shown to govern whether or not the first passage time distribution is well represented by its mean. The classical random walk problem of return to the origin (recurrence) was shown equivalent to the Landau-Peierls instability of elastic networks thus closing a the circle with our starting point.

9.4 Publications

In what follows articles appear as published in the scientific press and according to the order at which they were discussed above.

1. S. Reuveni, R. Granek and J. Klafter, Proteins: coexistence of stability and flexibility. *Phys. Rev. Lett.* **100**, 208101, (2008).

2. M. de Leeuw, S. Reuveni, J. Klafter and R. Granek, Coexistence of flexibility and stability of proteins: an equation of state. *PLoS ONE,* **4**(10), (2009).

3. S. Reuveni, R. Granek and J. Klafter, Anomalies in the vibrational dynamics of proteins are a consequence of fractal like structure. *Proc. Natl. Acad. Sci.,* **107**, 13696, (2010).

4. S. Reuveni, J. Klafter and R. Granek, Dynamic Structure Factor of Vibrating Fractals. *Phys. Rev. Lett.,* **108**, 068101, (2012).

5. S. Reuveni, J. Klafter and R. Granek, Dynamic Structure Factor of Vibrating Fractals: Proteins as a Case Study. *Phys. Rev. E.,* **85**, 011906, (2012).

6. S. Reuveni, R. Granek and J. Klafter, Vibrational shortcut to the mean-first-passage-time problem. *Phys. Rev. E.,* **81**, 040103(R), (2010).

7. S. Reuveni, R. Granek and J. Klafter, General Mapping Between Random Walk and Thermal Vibrations in Elastic Networks: Fractal Networks as a Case Study. *Phys. Rev. E.,* **82**, 041132, (2010).

Proteins: Coexistence of Stability and Flexibility

Shlomi Reuveni,[1] Rony Granek,[2] and Joseph Klafter[1]

[1]School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel

[2]Department of Biotechnology Engineering, Ben-Gurion University, Beer Sheva 84105, Israel

(Received 6 March 2008; published 19 May 2008)

We introduce an equation for protein native topology based on recent analysis of data from the Protein Data Bank and on a generalization of the Landau-Peierls instability criterion for fractals. The equation relates the protein fractal dimension d_f, the spectral dimension d_s, and the number of amino acids N. Deviations from the equation may render a protein unfolded. The fractal nature of proteins is shown to bridge their seemingly conflicting properties of stability and flexibility. Over 500 proteins have been analyzed (d_f, d_s, and N) and found to obey this equation of state.

DOI: 10.1103/PhysRevLett.100.208101 PACS numbers: 87.14.E−, 05.40.−a, 05.45.Df, 87.15.hp

Two seemingly conflicting properties of native proteins, such as enzymes and antibodies, are known to coexist. While proteins need to keep their specific native fold structure thermally stable, the native fold displays the ability to perform flexible motions that allow proper function [1–4]. This conflict cannot be bridged by compact objects which are characterized by small amplitude vibrations and by a Debye density of low frequency modes. Recently, however, it became clear that proteins can be described as fractals: namely, geometrical objects that possess self-similarity [5–9]. Adopting the fractal point of view to proteins makes it possible to describe within the same framework essential information regarding topology and dynamics [10,11] using three parameters: the number of amino acids along the protein backbone N, the spectral dimension d_s, and the fractal dimension d_f.

Based on a generalization [12] of the Landau-Peierls instability criterion [13], we derive a relation between the spectral dimension d_s, the fractal dimension d_f, and the number of amino acids along the protein backbone:

$$\frac{2}{d_s} + \frac{1}{d_f} = 1 + \frac{b}{\ln(N)}. \tag{1}$$

The spectral dimension d_s governs the density of low frequency normal modes of a fractal or protein. More precisely, denoting the density of modes $g(\omega)$, the scaling relation $g(\omega) \sim \omega^{d_s - 1}$ holds for low frequencies. Describing the mass fractal dimension d_f is most convenient using a three-dimensional example. Draw a sphere of radius r enclosing some lattice points in space and calculate their mass $M(r)$, increase r and calculate again. Do this several times and if $M(r)$ scales as r^{d_f} the exponent d_f is called the fractal dimension. For a regular 3D lattice both d_s and d_f coincide with the usual dimension of 3. For proteins, however, it is usually found that $d_s < 2$ and $2 < d_f < 3$, leading to an excess of low frequency modes and a more sparse fill of space [5,6,9,14,15]. The parameter b weakly depends on temperature and interaction parameters as discussed later.

Equation (1) is obeyed by a large class of proteins regardless of their source or function. It should be noted that every protein in our set, bearing an entirely different sequence, is in fact a *different* physical system. Thus Eq. (1) describes the universal common behavior of these different systems, as opposed to the case of a single underlying system, e.g., a Gaussian or swollen polymer chain, studied at many different sizes. We suggest that deviations from this ''equation of state'' for protein topology may render a protein unfolded. The fractal character implies large amplitude vibrations of the protein that could have led to unfolding. By selecting a thermodynamic state that is ''close'' to the edge of stability against unfolding, nature has solved the thermostability conflict. Nature's solution might be incorporated when planning biologically inspired catalysts.

We are led to relation (1) from two different independent pathways. The first approach utilizes the Gaussian network model (GNM) [16]. The melting of a protein is treated in this approach in a way similar to the melting of a solid crystal [17], with an additional assumption: surface residues initiate the melting process in proteins. Another approach that leads to relation (1) is motivated by the viewpoint of a folded protein as a collapsed polymer. It introduces a non-Lindemann criterion and a bond-bending Hamiltonian rather than the GNM Hamiltonian used in the first approach.

The GNM considers proteins to be elastic networks whose nodes correspond to the positions of the α carbons in the native structure, and the interactions among nodes are modeled as homogeneous harmonic springs. An interaction between two nodes exists only if the nodes are separated by a distance less than R_c, a distance known as the interaction cutoff. The cutoff distance is usually taken in the range 6–7 Å, based on the radius of the first coordination shell around residues observed in the Protein Data Bank (PDB) structures [18,19]. The GNM is defined by the

harmonic potential energy:

$$V_{\mathrm{GNM}} = \frac{\gamma}{2}\sum_{i,j}\Delta_{ij}(\delta\vec{r}_i - \delta\vec{r}_j)^2. \quad (2)$$

Here γ is the springs force constant and is assumed to be homogeneous and $\delta\vec{r}_i$ is the displacement with respect to the equilibrium position $\vec{R}_i^0$ of the ith C_α atom. Δ_{ij} is the network connectivity matrix with the following entries: $\Delta_{ij} = 1$ if $i \neq j$ and the distance $|\vec{R}_i^0 - \vec{R}_j^0|$ between two C_α atoms in the native conformation is smaller than R_c, $\Delta_{ij} = 0$ otherwise. The spectrum of the elastic network is given by the set of eigenvalues $\{\omega_0^2, \omega_1^2, \ldots, \omega_{N-1}^2\}$ of the Kirchhoff matrix, $\Gamma_{ij} = -\Delta_{ij} + \delta_{ij}\sum_{k\neq i}\Delta_{ik}$. The only information required to implement the method is the knowledge of the native structure. GNM has been widely applied because it yields results in agreement with x-ray spectroscopy and NMR experiments [16,20].

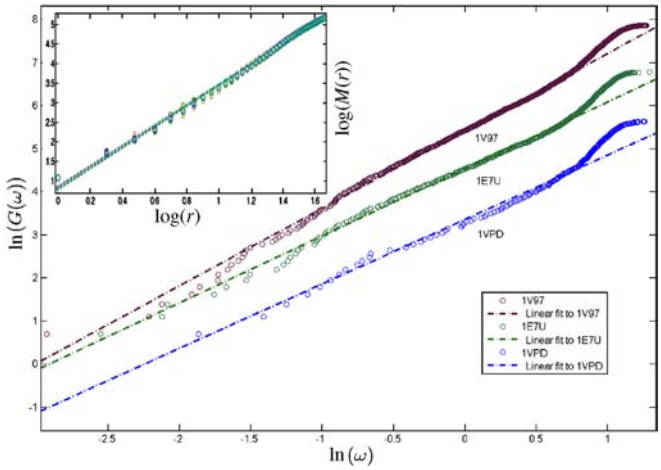


FIG. 1 (color online). Calculating the spectral dimension d_s for PDB codes: 1V97 ($N = 2594$, $d_s = 1.78$), 1E7U ($N = 872$, $d_s = 1.56$), and 1VPD ($N = 279$, $d_s = 1.49$). For each protein, we found the set of vibrational eigenfrequencies $\{\omega_0, \omega_1, \ldots, \omega_{N-1}\}$ that characterize the elastic network it forms when modeled by the GNM and plotted $\ln[G(\omega)]$ vs $\ln(\omega)$. In this example $R_c = 6$ Å and $G(\omega)$ is the cumulative density of modes defined as $G(\omega) = \int_0^\omega g(\omega')d\omega'$. All obtained modes are shown. The low frequency regions of $G(\omega)$ clearly exhibit a power law behavior; i.e., the scaling relation $G(\omega) \sim \omega^{d_s}$ holds for low frequencies. Dashed lines indicate best fits to these regions, $\omega \in [0.109, 1.67]$, $\omega \in [0.119, 1.6]$, and $\omega \in [0.243, 0.888]$ for 1V97, 1E7U, and 1VPD, correspondingly; the slopes correspond to the spectral dimensions. Inset: Calculating the mass fractal dimension d_f for PDB code 1V97 ($N = 2594$, $d_f = 2.64$), d_f was taken to be the average mass fractal dimension obtained by choosing the origin to be each and every one of the ten C-α atoms closest to the protein's center of mass. For a given origin, d_f was estimated via a power law fitting to $M(r)$, dashed lines indicate best fits. The data points as well as the best fits for different origins overlap significantly; we take the average slope to be the fractal dimension.

In order to test the validity of Eq. (1), we calculated the spectral and fractal dimensions for a data set of 543 proteins; see Fig. 1. Calculations were preformed on known protein structures, all structures were downloaded from the PDB [21]. The proteins that were chosen may differ in function and/or source organism and represent a wide length scale ranging from 100 to 3000 residues. Statistical analysis of the data gathered reveals satisfying agreement with Eq. (1). Fitting our data with Eq. (1) yields the following best-fit parameters: $b = 2.80$ for the cutoff $R_c = 7$ Å and $b = 3.97$ for a slightly different cutoff $R_c = 6$ Å. Despite the diversity in the sample data both cases yield significant correlation coefficients: 0.64 for $R_c = 7$ Å and 0.55 for $R_c = 6$ Å. In what follows we use the latter cutoff. Testing the validity of our predictions further, we tried fitting the data with the equation $\frac{2}{d_s} + \frac{1}{d_f} = a + \frac{b}{\ln(N)}$, which is a modification of Eq. (1) with the unity replaced by a parameter "a" on the right-hand side. The results are shown in Fig. 2. Allowing a free constant fitting parameter enabled us to confront theory with practice since our prediction is $a = 1$. A similar relation between N and d_s was suggested and tested on a small set of proteins in [6]. In that study a peculiar offset in the observed value of a constant fitting parameter, predicted to be exactly unity, was reported. We believe to have explained the reason for this offset and by doing so we were led to Eq. (1). The results shown in Fig. 2 indicate that the value of the parameter "a" is indeed close to 1. We also checked the validity of Eq. (1) for proteins all originating from the same creature. We thus "sliced" the data according to various sources (human, *E. coli*, etc., ...) in order to gain further insight into the relation between the source organism and the fitting parameters. The results of this analysis are summarized in Table I. Of special interest are proteins

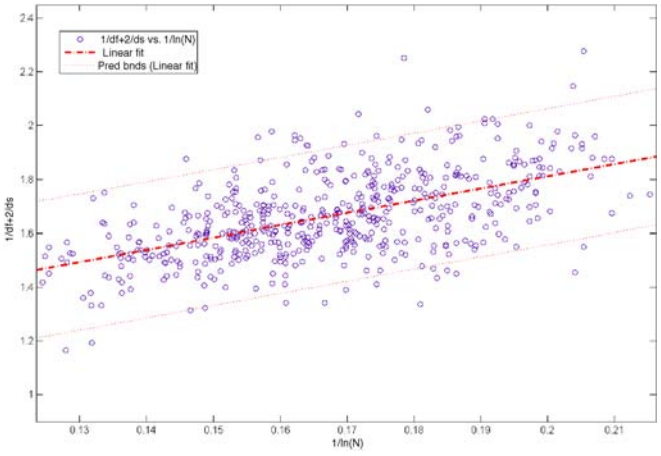


FIG. 2 (color online). Fitting the data gathered for 543 proteins with the equation $\frac{2}{d_s} + \frac{1}{d_f} = a + \frac{b}{\ln(N)}$. Here, the spectral dimension was calculated for $R_c = 6$ Å. The best-fit parameters are $a = 0.90$ and $b = 4.53$, the correlation coefficient is 0.55. Prediction bounds are for a confidence level of 95%.

TABLE I. Fitting the data from various creatures with the equation $\frac{2}{d_s} + \frac{1}{d_f} = a + \frac{b}{\ln(N)}$. Here, the spectral dimension was calculated for $R_c = 6$ Å, and c.c is the correlation coefficient. It is apparent from the table that when allowing a constant fitting parameter its value remains close to 1; this is true for both the set as a whole and for the overwhelming majority of creatures we analyzed.

Source	Proteins	a	b	c.c.
All	543	0.9 ± 0.09	4.53 ± 0.57	0.55
Mesophiles	432	0.91 ± 0.1	4.45 ± 0.61	0.57
E. coli	40	1.05 ± 0.25	3.66 ± 1.53	0.62
Bacillus subtilis	40	0.66 ± 0.42	6.01 ± 2.49	0.62
Bos taurus (cattle)	36	1 ± 0.30	3.71 ± 1.79	0.59
Homo sapiens (human)	44	1.13 ± 0.43	3.21 ± 2.54	0.36
Mus musculus (mouse)	37	1.18 ± 0.40	3.11 ± 2.26	0.43
Rattus norvegicus (rat)	36	0.86 ± 0.47	5.12 ± 2.71	0.55
Saccharomyces cerevisiae (yeast)	38	0.81 ± 0.47	5.05 ± 3.08	0.55
Salmonella typhimurium	28	0.59 ± 0.50	6.45 ± 3.08	0.64
Hyperthermophiles	111	0.87 ± 0.25	4.8 ± 1.52	0.51
Pyrococcus	44	0.99 ± 0.42	3.9 ± 2.57	0.42
T. maritima	49	0.95 ± 0.46	4.46 ± 2.70	0.44
A. aeolicus	20	0.73 ± 0.40	5.84 ± 2.36	0.77

originating in hyperthermophiles [22]. Surprisingly, such proteins that were included in the analyzed data, Fig. 2 and Table I, appear to fulfill Eq. (1).

We now describe the physics behind Eq. (1) and the alternative routes leading to it. In a paper generalizing the Landau-Peierls instability, Burioni *et al.* [12] showed that for $d_s < 2$ the mean square displacement (MSD) $\langle \delta r^2 \rangle$ of a structural unit (in the GNM, a single amino acid) in a system composed of N elements diverges in the limit $N \to \infty$ as

$$\langle \delta r^2 \rangle \propto \frac{k_B T}{\gamma} N^{(2/d_s)-1}. \tag{3}$$

It is clear that with $d_s < 2$, $\langle \delta r^2 \rangle$ grows indefinitely with N. Letting p be the ratio between the number of surface residues and the total number of residues in a protein and $q = 1 - p$ we write $\langle \delta r^2 \rangle_{\text{total}} = p \langle \delta r^2 \rangle_{\text{surface}} + q \langle \delta r^2 \rangle_{\text{bulk}}$. For this equation to hold for every N, both terms on the right-hand side must scale as the left-hand side, i.e., as in Eq. (3). Since by definition p is directly proportional to the surface to volume ratio of a protein, we obtain

$$p \propto \frac{S}{V} \propto \frac{1}{R_g} \propto \frac{1}{N^{1/d_f}}, \tag{4}$$

where R_g is the gyration radius of the protein [5,23]. At very low temperatures the MSDs of surface residues and of bulk residues are of the same order of magnitude. As temperature increases, the MSD values grow, and since surface residues are those prone to interactions with the solvent, it is reasonable to assume that melting starts when MSD values of surface residues reach a certain threshold. Denoting this threshold $\langle \delta r^2_{\text{melting}} \rangle_{\text{surface}}$, letting T_m represent the melting temperature and utilizing the scaling law of $p \langle \delta r^2 \rangle_{\text{surface}}$, we obtain the following approximation:

$$\frac{k_B T_m}{\gamma} N^{(2/d_s)+(1/d_f)-1} \propto \langle \delta r^2_{\text{melting}} \rangle_{\text{surface}}. \tag{5}$$

Rearrangement leads to Eq. (1), where the constant b depends on the parameters $\langle \delta r^2_{\text{melting}} \rangle_{\text{surface}}$, γ, and T_m. This dependence, however, is logarithmic and thus very weak, allowing a comparison among different proteins without computation of the specific parameters.

A different route to Eq. (1) is to start with a tensorial elasticity model rather than the scalar elasticity (Born) model described by the GNM. Here we use the bond-bending potential, previously studied for percolation [24,25]:

$$V = \frac{\gamma}{2} \sum_{ij} \Delta_{ij} [(\delta \vec{r}_i - \delta \vec{r}_j) \cdot \hat{r}_{ij}]^2 + \frac{B}{2} \sum_{jik} \Delta_{ij} \Delta_{ik} (\delta \theta_{jik})^2, \tag{6}$$

where $\delta \theta_{jik}$ is the angle between bonds $\langle ij \rangle$ and $\langle ik \rangle$, and $\hat{r}_{ij}$ is the unit vector along the bond $\langle ij \rangle$. We note that the first term is essentially the anisotropic network model discussed by Atilgan and co-workers [26] and describes the stretch-compress penalty, and the second term describes bond-bending penalty. When the bond-bending potentials are effectively softer than stretch-compress potentials ($B \ll \gamma R_g^2$), a very likely situation in proteins, the density of low frequency modes is dominated by bond-bending behavior and $g(\omega) \sim \omega^{d_E - 1}$, where d_E is the bond-bending spectral dimension equivalent to the spectral dimension d_s. For percolation clusters $d_E < 1$, and this is expected also for other fractals.

Next consider the variance of fluctuations in the distance between two tagged points on the protein that are distanced R_g apart. This may be evaluated in a similar way to the one described in [10,27], as $\langle \vec{x}^2(R_g)\rangle \sim N^{(2/d_E)-1}$. Importantly, if $d_E < 1$ and $d_f > 2$, this diverges with increasing N faster than $R_g^2 \sim N^{2/d_f}$. We postulate that melting occurs when the magnitude of these fluctuations reaches the protein size, namely, when $\langle \vec{x}^2(R_g)\rangle \sim R_g^2$. This leads to

$$\frac{2}{d_E} - 1 - \frac{2}{d_f} = \frac{\text{const}}{\ln(N)}, \tag{7}$$

an equation that resembles Eq. (1) with d_s replaced by d_E.

In order to find d_E one has to solve for the eigenfrequencies of the bond-bending Hamiltonian. To circumvent this difficulty, we use relations that have been derived for percolation clusters, assuming that they hold for other fractals and therefore at least approximately for protein networks [24,25]. The spectral and bond-bending spectral dimensions have been shown to obey the relations [24,25,28] $d_s = \frac{2d_f}{2-d+d_f+t/\nu}$ and $d_E = \frac{2d_f}{d_f+2+1/\nu}$, where t is the percolation conductivity exponent $\sigma \sim (p - p_c)^t$ and ν is the percolation correlation length exponent $\xi \sim (p - p_c)^{-\nu}$. From these two relations we find, for $d = 3$,

$$\frac{2}{d_E} - 1 - \frac{2}{d_f} \propto \frac{2}{d_s} - 1 + \frac{1}{d_f}. \tag{8}$$

Using Eqs. (7) and (8) leads again to Eq. (1). The relation $\frac{2}{d_s} + \frac{1}{d_f} = 1 + \frac{b}{\ln(N)}$ and the general inequalities $1 \le d_s \le d_f \le 3$ lead to the following effective bounds on d_s and d_f: $1 \le d_s \le \frac{3}{1+b/\ln N} \le d_f \le 3$. Interestingly, the latter bounds permit values of d_s greater than 2. This does not pose any conflict since the Landau-Peierls instability is controlled in this bond-bending model by d_E rather than d_s.

One may wonder what will happen if a protein is forced to strongly deviate from Eq. (1) and how artificial deformations of the protein fold may lead to a breakdown of criterion (1). Strong deformations of the protein fold may actually happen *in vivo* as part of a natural process. A possible example is GroEL, a protein chaperon that is required for the proper folding of many proteins. Recent molecular dynamics simulations demonstrate the unfolding action of GroEL on a protein substrate [29,30]. Our work provides a theoretical framework that may help understand GroEL induced unfolding. In addition our work opens new possibilities for nanoscale and biologically inspired engineering of catalysts, emphasizing the importance of internal motion.

We thank Tal Shapiro and Noam Grimberg for assistance with data analysis. R. G. and J. K. acknowledge support from The Israel Science Foundation. R. G. is a member of the Ilse Katz Center for Meso and Nanoscale Science and Technology and the Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering.

[1] D. Joseph, G. A. Petsko, and M. Karplus, Science **249**, 1425 (1990).
[2] P. J. Steinbach *et al.*, Biochemistry **30**, 3988 (1991).
[3] B. F. Rasmussen, A. M. Stock, D. Ringe, and G. A. Petsko, Nature (London) **357**, 423 (1992).
[4] A. Henzler-Wildman *et al.*, Nature (London) **450**, 913 (2007).
[5] M. B. Enright and D. M. Leitner, Phys. Rev. E **71**, 011912 (2005).
[6] R. Burioni, D. Cassi, F. Cecconi, and A. Vulpiani, Proteins: Struct. Funct. Genet. **55**, 529 (2004).
[7] S. Lushnikov, A. Svanidze, and I. Sashin, J. Exp. Theor. Phys. Lett. **82**, 30 (2005).
[8] H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinson, and S. R. Kurtz, Phys. Rev. Lett. **45**, 1456 (1980).
[9] R. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).
[10] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[11] Nature (London) **437**, 172 (2005).
[12] R. Burioni, D. Cassi, M. P. Fontana, and A. Vulpiani, Europhys. Lett. **58**, 806 (2002).
[13] R. Peierls, Helv. Phys. Acta **7**, Suppl. 2, 81 (1934).
[14] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).
[15] S. Ciliberti, P. DeLosRios, and F. Piazza, Phys. Rev. Lett. **96**, 198103 (2006).
[16] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[17] Y. Zhou, M. Karplus, K. D. Ball, and R. S. Berry, J. Chem. Phys. **116**, 2323 (2002).
[18] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[19] I. Bahar and R. L. Jernigan, J. Mol. Biol. **266**, 195 (1997).
[20] L.-W. Yang, E. Eyal, C. Chennubhotla, J. JunGoo, A. M. Gronenborn, and I. Bahar, Structure **15**, 741 (2007).
[21] F. C. Bernstein *et al.*, J. Mol. Biol. **112**, 535 (1977).
[22] I. N. Berezovsky and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **102**, 12 742 (2005).
[23] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University, Ithaca, New York, 1979).
[24] S. Feng, Phys. Rev. B **32**, 5793 (1985).
[25] I. Webman and G. S. Grest, Phys. Rev. B **31**, 1689 (1985).
[26] A. R. Atilgan, S. R. Durrell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[27] S. Alexander, Phys. Rev. B **40**, 7953 (1989).
[28] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992), 2nd ed.
[29] A. Van der Vaart, J. Ma, and M. Karplus, Biophys. J. **87**, 562 (2004).
[30] G. Stan, G. H. Lorimer, D. Thirumalai, and B. R. Brooks, Proc. Natl. Acad. Sci. U.S.A. **104**, 8803 (2007).

OPEN ACCESS Freely available online

PLoS one

Coexistence of Flexibility and Stability of Proteins: An Equation of State

Marina de Leeuw[1], Shlomi Reuveni[2], Joseph Klafter[2,3], Rony Granek[1]*

1 Department of Biotechnology Engineering, Ben-Gurion University of The Negev, Beer Sheva, Israel, **2** School of Chemistry, Tel-Aviv University, Tel-Aviv, Israel, **3** Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwig-Universität Freiburg, Freiburg, Germany

Abstract

We consider a recently suggested "equation of state" for natively folded proteins, and verify its validity for a set of about 5800 proteins. The equation is based on a fractal viewpoint of proteins, on a generalization of the Landau-Peierls instability, and on a marginal stability criterion. The latter allows for coexistence of stability and flexibility of proteins, which is required for their proper function. The equation of state relates the protein fractal dimension d_f, its spectral dimension d_s, and the number of amino acids *N*. Using structural data from the protein data bank (PDB) and the Gaussian network model (GNM), we compute d_f and d_s for the entire set and demonstrate that the equation of state is well obeyed. Addressing the fractal properties and making use of the equation of state may help to engineer biologically inspired catalysts.

Citation: de Leeuw M, Reuveni S, Klafter J, Granek R (2009) Coexistence of Flexibility and Stability of Proteins: An Equation of State. PLoS ONE 4(10): e7296. doi:10.1371/journal.pone.0007296

Editor: Roland Dunbrack, Fox Chase Cancer Center, United States of America

Received July 15, 2009; **Accepted** August 24, 2009; **Published** October 9, 2009

Funding: Rony Granek and Joseph Klafter thank The Israel Science Foundation for financial support. Joseph Klafter acknowledges support from the Excellence Initiative of the German Federal and State Governments. Shlomi Reuveni acknowledges support of the Converging Technologies fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rgranek@bgu.ac.il

Introduction

Proteins are one of the major components of living cells. They constitute more than half of the cell's dry weight, and are responsible for the execution of most cellular functions required for life, including among others, catalysis and molecular recognition within and between cells and their surroundings. Understanding the relationships between structure, internal dynamics, and enzymatic activity at the single-molecule level could pave new ways to manipulate individual molecules.

Two seemingly conflicting properties of native proteins, such as enzymes and antibodies, are known to coexist. While proteins need to keep their specific native fold structure thermally stable, the native fold displays the ability to perform large amplitude conformational changes that allow proper function [1]. This conflict cannot be bridged by compact objects which are characterized by small amplitude vibrations [2]. Recently, however, it became evident that proteins can be described as fractals; namely, geometrical objects that possess self similarity [3,4]. Adopting the fractal point of view to proteins makes it possible to describe within the same framework essential information regarding topology and dynamics.

Based on the fractal viewpoint, we have recently derived a universal equation of state for protein topology. The same fractal viewpoint allows describing the near equilibrium dynamics of native proteins. We have recently shown that it leads to anomalous dynamics [5]. For example, the autocorrelation function of the distance between two α-carbons on a protein is predicted to decay anomalously, first, at short times, as $1-t^{\delta}$ and later, at long times, as $t^{-\beta}$, where δ and β are exponents that depend on various fractal dimensions. This type of relaxation has been recently observed in single molecule experiments [6,7]. Closely related is the anomalous diffusion of an α-carbon that is predicted by the fractal model, where the mean square displacement is found to increase as $\sim t^{\delta}$. Such a behavior has also been recently observed in molecular dynamics simulations [8].

Natively folded proteins can be characterized by broken dimensions: the fractal and spectral dimensions [2,4,5,9–12]. The mass fractal dimension d_f describes the spatial distribution of the mass within the protein via the scaling relation $M(r)\sim r^{d_f}$, where $M(r)$ is the mass enclosed in a sphere of radius r [3]. The spectral dimension d_s governs the density of low frequency vibrational normal modes *via* the scaling relation $g(\omega)\sim\omega^{d_s-1}$, where $g(\omega)d\omega$ is the number of modes in the frequency range $[\omega,\omega+d\omega]$ [13]. While for regular three dimensional (3D) lattices both d_s and d_f coincide with the usual dimension of 3, for proteins it is usually found that $d_s<2$ and $2<d_f<3$, leading to an excess of low frequency modes and a more sparse fill of space [2,4,12]. Importantly, the regime $d_s<2$ is associated with the so-called Landau-Peierls instability, where the amplitude of vibrations increases with the number of residues N [14,15]. As this amplitude overcomes a threshold value, it may cause the protein to unfold [2,12].

The Landau-Peirels instability is most readily derived using the density of states. The static mean square displacement (MSD) of an α-carbon, which is essentially the so-called B_i-factor, averaged over all α-carbons of the protein, may be expressed as

$$\langle(\Delta u_i)^2\rangle=\frac{k_BT}{m}\int_{\omega_{\min}}^{\omega_{\max}}d\omega\,\frac{g(\omega)}{\omega^2} \qquad (1)$$

where m is the average mass of an amino acid. Since $g(\omega)\sim\omega^{d_s-1}$, it follows that if $d_s<2$ the integral diverges with the lower bound

$\omega_{\min}$. The latter depends on the protein radius of gyration R_g and the number of residues N as $\omega_{\min} \sim {R_g}^{-d_f/d_s} \sim N^{-1/d_s}$. This leads to $\langle(\Delta u_i)^2\rangle \sim N^{2/d_s-1}$, which increases with N for $d_s<2$. In particular, the static MSD of *surface* residues has been argued to grow as

$$\langle(\Delta u_i)^2\rangle_{surface} \sim N^{2/d_s+1/d_f-1} \quad (2)$$

We have proposed a marginal stability criterion [16], in which most proteins "exploit" the Landau-Peierls instability to attain large amplitude vibrations, which is required for their proper function, yet maintaining their native fold. Thus proteins are assumed to exist in a thermodynamic state close to the edge of unfolding. Based on this and the Landau-Peirels instability of the surface residues, Eq. (2), a general equation of state has been proposed that relates between the spectral dimension d_s, the fractal dimension d_f, and the number of amino acids along the protein backbone N:

$$\frac{2}{d_s}+\frac{1}{d_f}=1+\frac{b}{\ln(N)} \quad (3)$$

where b is a molecular fit parameter depending on the temperature T, the GNM spring constant γ, and the GNM cutoff R_c: $b \approx \ln(\gamma {R_c}^2/k_B T)$ [12]. It has been shown that this equation is obeyed by about 500 proteins regardless of their source or function [12]. In the present study we check the validity of Eq. (3) for a much larger set of over 5,000 proteins, using a range of statistical methods, and show that also for this very large set Eq. (3) is beautifully fulfilled. This supports the marginal stability criterion that led to this equation.

Methods

We have used all data files present in the Protein Data Bank (PDB) [17] and filtered out proteins exceeding 95% sequence identity and proteins that have ligands, RNA, or DNA. We have also removed incomplete data files, files that contained data of the α-carbons alone, and also files of proteins smaller than 100 amino acids that are too small to be characterized as fractals. With this screening the set has been reduced to 5793 proteins.

The fractal and spectral dimensions were calculated for all 5793 proteins in similar ways to the procedure described by [12]. Finding the protein center of mass and placing the origin of coordinates at the ten nearest α-carbons, the mass was calculated as a function of the distance r on a log-log scale. The fractal dimension d_f has been obtained as the slope of this plot for distances below the protein gyration radius R_g, averaged over the ten origin of coordinates, see examples in Fig. 1. It should be noted that when a few alternative locations of an atom are given, only the "*A*" location (usually the most abundant one) has been used.

To find the spectral dimension d_s, we calculate the cumulative density of normal vibrational modes $G(\omega)$, $G(\omega)=\int_0^{\omega} d\omega' g(\omega')$, representing the number of modes up to a frequency ω. To obtain the vibrational modes, we used a frequently applied elastic model for protein vibrations, the Gaussian network model (GNM) [12,18–23]. Two values were taken for the interaction distance cutoff R_c, that describes the range of the interaction between an α-carbons pair, $R_c=6$ Å and $R_c=7$ Å. Plotting on a log-log scale $G(\omega)$ against the frequency ω, the slope in the low frequency range (containing about 24% of the modes, independent of the protein type or size N) defines d_s, i.e. $G(\omega) \sim \omega^{d_s}$, see examples in Fig. 2 for the case $R_c=6$ Å.

To deal with the large number of proteins in this set, both procedures were automated using suitable computer codes. The

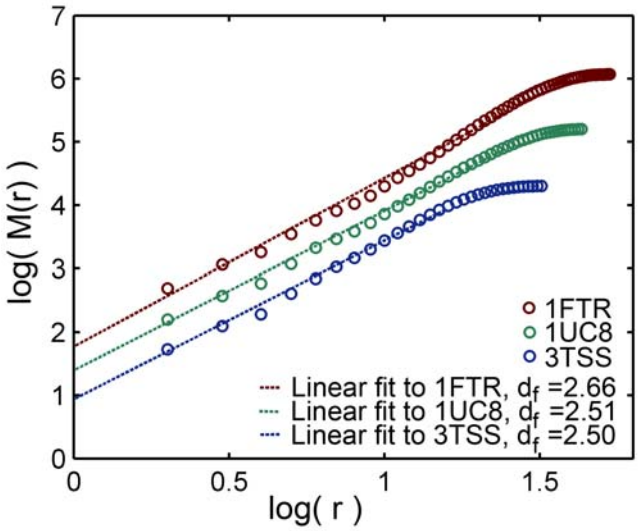


Figure 1. Fractal dimension. The fractal dimension of three selected proteins: 1FTR (1184 amino acids, $d_f=2.66$), 1UC8 (505 amino acids, $d_f=2.51$) and 3TSS (190 amino acids, $d_f=2.50$). The mass $M(r)$ enclosed in concentric spheres of radius r is plotted against r (measured in units of Å) on a log-log scale and the slope determines the fractal dimension, $M(r) \sim r^{d_f}$. The plots of 1FTR and 1UC8 were shifted along the y axis (+1 and +0.5 respectively) for clarity.
doi:10.1371/journal.pone.0007296.g001

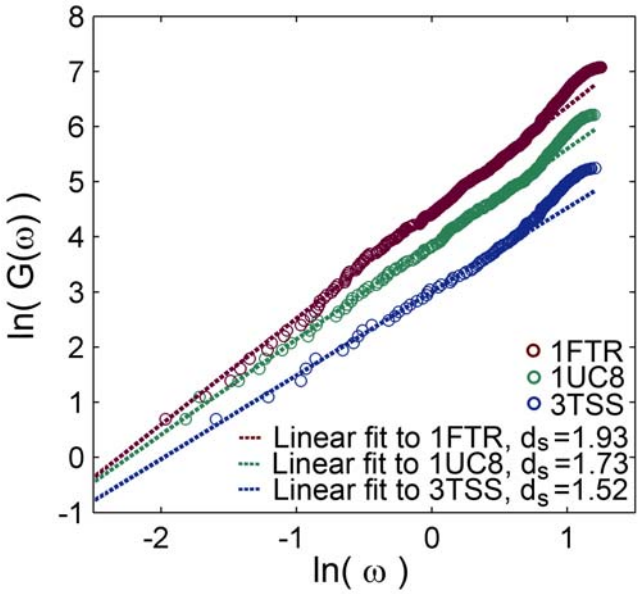


Figure 2. Spectral dimension. The spectral dimension of three selected proteins (same proteins as in Fig. 1): 1FTR (1184 amino acids, $d_s=1.93$), 1UC8 (505 amino acids, $d_s=1.73$) and 3TSS (190 amino acids, $d_s=1.52$). The cumulative density of normal modes $G(\omega)$ is plotted against the frequency ω (measured in units of the spring natural frequency) on a log-log scale and the slope determines the spectral dimension, $G(\omega) \sim \omega^{d_s}$.
doi:10.1371/journal.pone.0007296.g002

automatically calculated spectral dimension values were compared (for the case $R_c = 7$ Å) to the manually obtained values for the set of 543 studied in [12]. We found almost vanishing mean of the difference between the two results (0.0034), showing that the error is statistical, and a low standard deviation (0.083), suggesting good agreement between the two methods of calculation.

In order to generally check for correlations between $2/d_s + 1/d_f$ and $1/\ln N$, simple regression was conducted (using SPSS). This shows statistical significance with $p<0.001$ and very high F-test values ($F(1,5791) = 3263$ and $F(1,4247) = 2120$ for $R_c = 6$ Å, $F(1,5791) = 4059$ and $F(1,3888) = 2314$ for $R_c = 7$ Å).

Results

The results for the whole set appear in the supporting information S1 and are shown in Figs. 3 and 4 (for $R_c = 6$ Å and 7 Å, respectively), where we plot the combination $2/d_s + 1/d_f$ against $1/\ln N$. In order to present the whole set of data, we designed a (smoothed) colored histogram based on a 100×100 grid, where a pixel color represent the number of proteins associated with the pixel. The data is first fitted to Eq. (3) (dashed lines). This leads to $b = 4.555$ for $R_c = 6$ Å (correlation coefficient $cc = 0.596$), see Fig. 3, and $b = 3.242$ for $R_c = 7$ Å ($cc = 0.605$), see Fig. 4. Using $b \approx \ln(\gamma R_c^2/k_B T)$, with $k_B T/\gamma$ in the range 0.5 Å^2 to 2 Å^2, we can estimate b to be in the range 3 to 5. The value of b is within the expected range.

We also fitted the data to an equation resembling Eq. (3) but in which the value "1" is replaced by a free parameter a:

$$\frac{2}{d_s} + \frac{1}{d_f} = a + \frac{b}{\ln(N)} \qquad (4)$$

This is done in order to verify if the free fit recovers the value $a = 1$. The results of this fit are also shown in Figs. 3–4 (full lines), and yield $a = 0.884$ and $b = 5.197$ for $R_c = 6$ Å (Fig. 3, $cc = 0.600$), and $a = 0.710$ and $b = 4.841$ for $R_c = 7$ Å (Fig. 4, $cc = 0.642$). Remarkably, the colored histogram shows a ridge roughly centered at the best fitting theoretical lines.

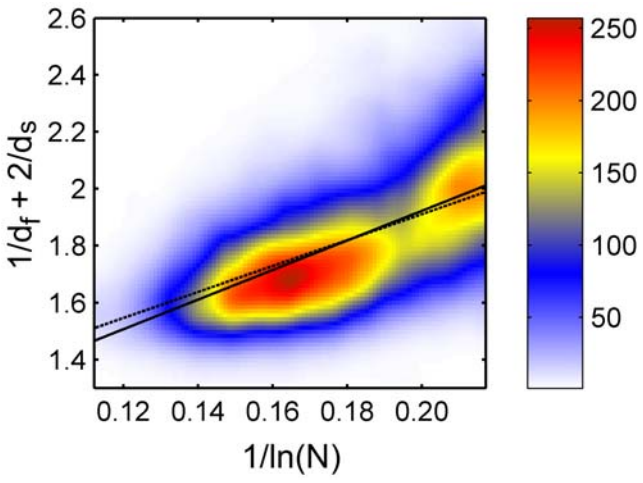


Figure 3. Full data set for $R_C = 6$ Å, colored histogram. The values of $2/d_s + 1/d_f$ against $1/\ln N$ plotted for the full data set (5793 proteins) with $R_C = 6$ Å. The data is presented using a smoothed colored histogram based on a 100×100 grid, see the color scale on the right (low density areas colored blue and high density red). The data was fitted to Eq. (3) (dashed line) and to Eq. (4) (full line).
doi:10.1371/journal.pone.0007296.g003

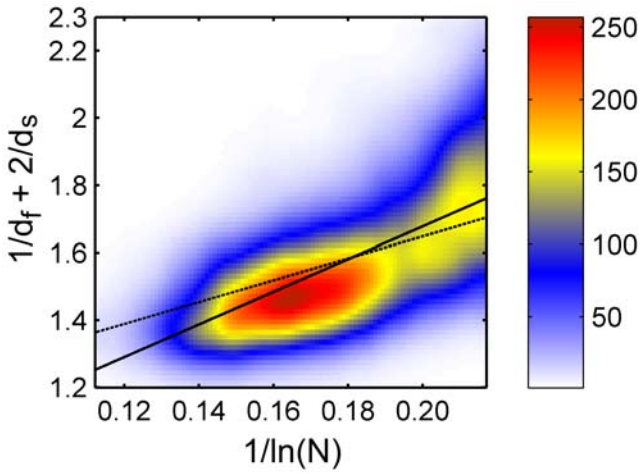


Figure 4. Full data set for $R_C = 7$ Å, colored histogram. Same as in Fig. 3 but for $R_C = 7$ Å.
doi:10.1371/journal.pone.0007296.g004

To improve the accuracy of the analyses, a subset was constructed containing only those proteins whose both d_f and d_s values have been determined with a *very high precision*, such that the squared correlation coefficients for the power-law fits of both $M(r)$ and $G(\omega)$ were in the range $R^2 > 0.99$. Accordingly, this subset for $R_c = 6$ Å (containing 4249 proteins) is not identical to the subset for $R_c = 7$ Å (containing 3890 proteins), see the supporting information S1 for details. The results are presented in Figs. 5–6. Fitting to Eq. (3) (dashed lines) leads to $b = 4.476$ for $R_c = 6$ Å (Fig. 5, $cc = 0.576$) and $b = 3.078$ for $R_c = 7$ Å (Fig. 6, $cc = 0.593$). Fitting the data to Eq. (4) (full lines), yields $a = 0.952$ and $b = 4.747$ for $R_c = 6$ Å (Fig. 5, $cc = 0.577$), and $a = 0.833$ and $b = 4.031$ for $R_c = 7$ Å (Fig. 6, $cc = 0.611$).

Although the data analysis presented in Fig. 3–6 appears complete, it fails to give equal weight to proteins of different sizes. All four different data sets used above are very rich in proteins of small (100–200 residues) and intermediate size, a consequence of

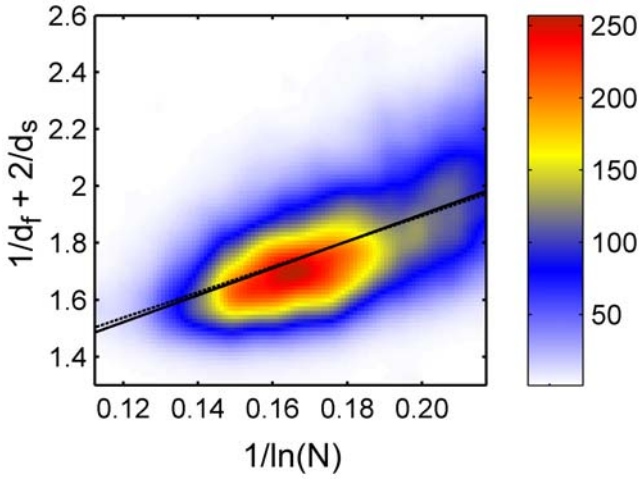


Figure 5. High precision data set for $R_C = 6$ Å, colored histogram. The values of $2/d_s + 1/d_f$ against $1/\ln N$ plotted for the refined subset of increased precision for $R_C = 6$ Å (4249 proteins), using a colored histogram (same as in Fig. 3).The data was fitted to Eq. (3) (dashed line) and to Eq. (4) (full line); the two lines are almost indistinguishable.
doi:10.1371/journal.pone.0007296.g005

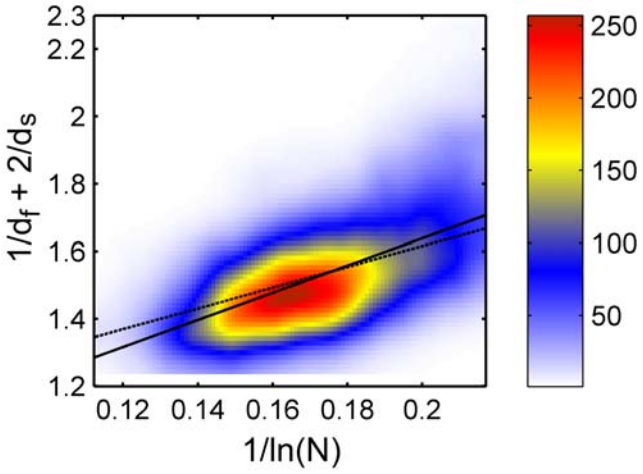


Figure 6. High precision data set for $R_C = 7$ Å, colored histogram. Same as in Fig. 5 but for the refined subset of increased precision for $R_C = 7$ Å (3890 proteins).
doi:10.1371/journal.pone.0007296.g006

their abundance in nature, while being poor in large proteins. Yet, the linear regression presented in Figs. 3–6 gives each protein an equal weight. Thus, while the small/intermediate size proteins are spread over a relatively limited range of N, they are overwhelming the linear regression, which is undesirable.

To circumvent this artifact, we have separated the x-axis $(1/\ln N)$ into 100 bins. For each bin we calculate the mean value of $2/d_s + 1/d_f$. The error of $2/d_s + 1/d_f$ for each bin is estimated as the standard deviation of this value. The results are summarized in Figs. 7, 8, 9, 10.

Results from the full set of 5793 proteins are presented in Figs. 7–8. Fitting to Eq. (3) (dashed lines) leads to b = 4.580 for $R_c = 6$ Å (Fig. 7, cc = 0.957) and b = 3.212 for $R_c = 7$ Å (Fig. 8, cc = 0.928). Fitting the data to Eq. (4) (full lines), yields a = 1.026 and b = 4.429 for $R_c = 6$ Å (Fig. 7, cc = 0.958), and a = 0.870 and b = 3.977 for $R_c = 7$ Å (Fig. 8, cc = 0.946). Note that all lines pass through almost all error bars, a remarkable result.

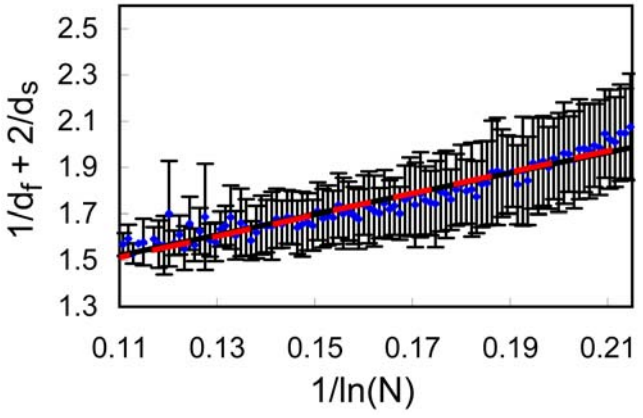


Figure 7. Full data set for $R_C = 6$ Å, division into bins. The values of $2/d_s + 1/d_f$ against $1/\ln N$ plotted for the full data set (5793 proteins) with $R_C = 6$ Å. The values of $1/\ln N$ were divided into 100 equally sized bins. For each bin we show the average value of $2/d_s + 1/d_f$ and the error bar presents its standard deviation. The data was fitted to Eq. (3)(dashed red line) and to Eq. (3) (full black line); the two lines are almost indistinguishable.
doi:10.1371/journal.pone.0007296.g007

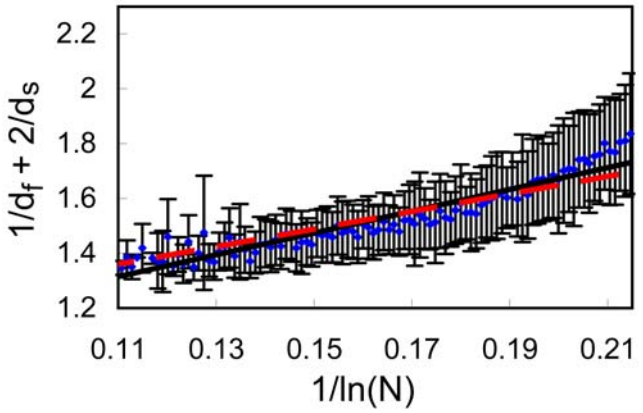


Figure 8. Full data set for $R_C = 7$ Å, division into bins. Same as in Fig. 7 but for $R_C = 7$ Å.
doi:10.1371/journal.pone.0007296.g008

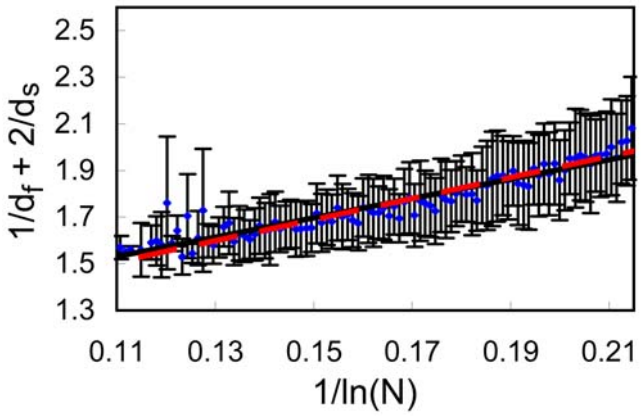


Figure 9. High precision data set for $R_C = 6$ Å, division into bins. Same as Fig. 7 but for the refined subset of increased precision for $R_C = 6$ Å (4249 proteins).
doi:10.1371/journal.pone.0007296.g009

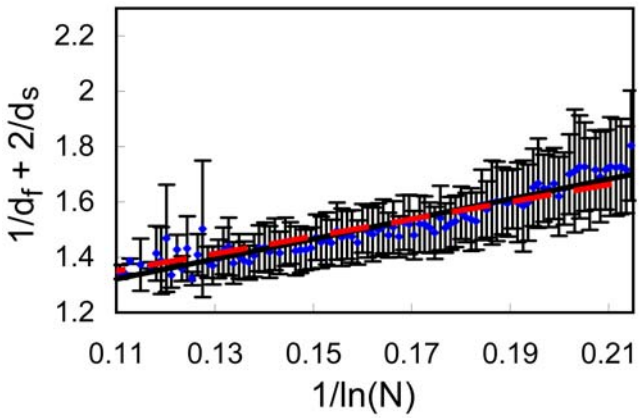


Figure 10. High precision data set for $R_C = 7$ Å, division into bins. Same as Fig. 7 but for the refined subset of increased precision for $R_C = 7$ Å (3890 proteins).
doi:10.1371/journal.pone.0007296.g010

In Figs. 9–10 we present results from the high precision subset of 4249 proteins. Fitting to Eq. (3) (dashed lines) leads to b = 4.535 for $R_c = 6$ Å (Fig. 9, cc = 0.941) and b = 3.124 for $R_c = 7$ Å (Fig. 10, cc = 0.937). Fitting the data to Eq. (4) (full lines), yields a = 1.065 and b = 4.155 for $R_c = 6$ Å (Fig. 9, cc = 0.945), and a = 0.917 and b = 3.609 for $R_c = 7$ Å (Fig. 10, cc = 0.946). Here, as well, all lines pass through almost error bars. This refined analysis gives an even stronger support to Eq. (3).

Discussion

All correlation coefficients mentioned above (Figs. 3–10) are considered excellent. In addition, the values of a are close to the theoretically predicted value $a = 1$, similar to the set of 543 proteins studied by [12]. In particular, the fits of the data to Eq. (4) for all data sets belonging to $R_c = 6$ Å (shown in Figs. 3,5,7 and 9) yields a values that are remarkably close to 1. The distribution of the data in all four sets appears as a ridge that is roughly centered at the best fitting theoretical lines (Figs. 3–6), and when the binning procedure is being used, all lines pass well through the error bars (Figs. 7–10). We believe that these results strongly confirm the universal behavior described by Eq. (3), thereby supporting the theoretical arguments leading to this equation.

Importantly, a is found to be particularly close to 1 when the binning procedure is introduced, in which we analyze the mean value of $2/d_s + 1/d_f$, for a given N, for its dependence on N. In these cases we also obtain remarkably good correlation coefficients, significantly better than those obtained without binning. This suggests that, as a group, proteins follow the equation of state, although the error bars indicate that there are other factors present that cause deviations from the equation. These factors could be related to the protein specific structure and/or function.

The distribution of the data in all four sets appears as a ridge that is roughly centered at the best fitting theoretical lines (Figs. 3–6), and when the binning procedure is being used, all lines pass well through the error bars (Figs. 7–10). We believe that these results strongly confirm the universal behavior described by Eq. (3), thereby supporting the theoretical arguments leading to this equation.

To conclude, our analysis confirms the fractal nature of proteins and supports the predicted universal equation of state (3). This suggests that the majority of proteins in the PDB exist in a marginally stable thermodynamic state, namely a state that is close to the edge of unfolding. This could be related to the fact that enzymes require flexibility and large internal motion to function properly [1]. We suggest that Eq. (3) can be used as a tool in the design of artificial enzymes [24]. Interestingly, fractal-like properties have also been suggested to appear in the configuration space of peptides [25].

Supporting Information

Supporting Information S1 A file containing the mass fractal dimension d_f and spectral dimension d_s of all proteins analyzed, divided into the four data sets described in the text: (i) GNM cutoff length 6 Å, (ii) GNM cutoff length 7 Å, (iii) GNM cutoff length 6 Å, a subset with high precision values of d_f and d_s, and (iv) GNM cutoff length 7 Å, a subset with high precision values of d_f and d_s. Found at: doi:10.1371/journal.pone.0007296.s001 (2.84 MB XLS)

Acknowledgments

We are grateful to Martin Karplus and Dave Thirumalai for illuminating discussions. RG is a member of the Ilse Katz Center for Meso and Nanoscale Science and Technology and the Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering.

Author Contributions

Analyzed the data: MdL. Wrote the paper: JK RG. Initiated and directed the research: RG. Contributed to data analysis: SR.

References

1. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, et al. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 450(7171): 913–916.
2. Burioni R, Cassi D, Cecconi F, Vulpiani A (2004) Topological thermal instability and length of proteins. Proteins-Structure Function and Bioinformatics 55(3): 529–535.
3. Stauffer D, Aharony A (1994) Introduction to percolation theory. ;CRC Press.
4. Enright MB, Leitner DM (2005) Mass fractal dimension and the compactness of proteins. Physical Review E 71(1): 011912.
5. Granek R, Klafter J (2005) Fractons in proteins: Can they lead to anomalously decaying time autocorrelations? Phys Rev Lett 95(9): 098106.
6. Kou SC, Xie XS (2004) Generalized langevin equation with fractional Gaussian noise: Subdiffusion within a single protein molecule. Phys Rev Lett 93(18): 180603.
7. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Observation of a power-law memory kernel for fluctuations within a single protein molecule. Phys Rev Lett 94(19): 198302.
8. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA (2008) How main-chains of proteins explore the free-energy landscape in native states. Proceedings of the National Academy of Sciences of the United States of America 105(50): 19708.
9. Stapleton HJ, Allen JP, Flynn CP, Stinson DG, Kurtz SR (1980) Fractal form of proteins. Phys Rev Lett 45(17): 1456–1459.
10. Elber R, Karplus M (1986) Low-frequency modes in proteins - Use of the effective-medium approximation to interpret the fractal dimension observed in electron-spin relaxation measurements. Phys Rev Lett 56(4): 394–397.
11. Lushnikov SG, Svanidze AV, Sashin IL (2005) Vibrational density of states of hen egg white lysozyme. Jetp Letters 82(1): 30–33.
12. Reuveni S, Granek R, Klafter J (2008) Proteins: Coexistence of stability and flexibility. Phys Rev Lett 100(20): 208101.
13. Alexander S (1989) Vibrations of fractals and scattering of light from aerogels. Physical Review B 40(11): 7953–7965.
14. Peierls R (1934) Bemerkungen über umwandlungstemperaturen. Helv.Phys.Acta 7: 81–83.
15. Burioni R, Cassi D, Fontana MP, Vulpiani A (2002) Vibrational thermodynamic instability of recursive networks. Europhys Lett 58(6): 806–810.
16. Li MS, Klimov DK, Thirumalai D (2004) Finite size effects on thermal denaturation of globular proteins. Phys Rev Lett 93(26): 268107.
17. Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, et al. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112(3): 535–542.
18. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2(3): 173–181.
19. Bahar I, Jernigan RL (1997) Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol 266(1): 195–214.
20. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. Phys Rev Lett 79(16): 3090–3093.
21. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80(1): 505–515.
22. Chennubhotla C, Rader AJ, Yang LW, Bahar I (2005) Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. Phys.Biol 2(4): S173–S180.
23. Yang LW, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, et al. (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure 15(6): 741–749.
24. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, et al. (2008) Kemp elimination catalysts by computational enzyme design. Nature 453(7192): 190–195.
25. Neusius T, Daidone I, Sokolov IM, Smith JC (2008) Subdiffusion in peptides originates from the fractal-like structure of configuration space. Phys Rev Lett 100(18): 188103.

Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure

Shlomi Reuveni[a,b,1], Rony Granek[c], and Joseph Klafter[a]

[a]School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel; [b]School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel; and [c]The Stella and Avram Goren-Goldstein Department of Biotechnology Engineering, Ben-Gurion University, Beer Sheva 84105, Israel

Edited* by Morrel H. Cohen, Rutgers, The State University of New Jersey, Bridgewater Township, NJ, and approved June 15, 2010 (received for review February 23, 2010)

Proteins have been shown to exhibit strange/anomalous dynamics displaying non-Debye density of vibrational states, anomalous spread of vibrational energy, large conformational changes, non-exponential decay of correlations, and nonexponential unfolding times. The anomalous behavior may, in principle, stem from various factors affecting the energy landscape under which a protein vibrates. Investigating the origins of such unconventional dynamics, we focus on the structure-dynamics interplay and introduce a stochastic approach to the vibrational dynamics of proteins. We use diffusion, a method sensitive to the structural features of the protein fold and them alone, in order to probe protein structure. Conducting a large-scale study of diffusion on over 500 Protein Data Bank structures we find it to be anomalous, an indication of a fractal-like structure. Taking advantage of known and newly derived relations between vibrational dynamics and diffusion, we demonstrate the equivalence of our findings to the existence of structurally originated anomalies in the vibrational dynamics of proteins. We conclude that these anomalies are a direct result of the fractal-like structure of proteins. The duality between diffusion and vibrational dynamics allows us to make, on a single-molecule level, experimentally testable predictions. The time dependent vibrational mean square displacement of an amino acid is predicted to be subdiffusive. The thermal variance in the instantaneous distance between amino acids is shown to grow as a power law of the equilibrium distance. Mean first passage time analysis is offered as a practical tool that may aid in the identification of amino acid pairs involved in large conformational changes.

protein dynamics | Gaussian network model | anomalous diffusion | intramolecular distance distributions | mean first passage time

Introduction

Proteins are commonly envisioned vibrating under the influence of a complicated energy landscape (1) held responsible for their intricate dynamics (1–9). Despite its overall complexity the energy landscape is approximately harmonic near its minima, a fact that does not seem to coincide with the rich dynamics mentioned above. Attacking the validity of the harmonic approximation is one way out of this puzzle. One might have suggested that the harmonic approximation is unable to capture the anomalous dynamics displayed by proteins since the latter is dominated by large excursions from the native state structure. The failure of the harmonic approximation then follows from the fact that it does not describe the energy of conformations which considerably deviate from the native state structure. In this paper we argue differently showing that the native state structure of proteins is fractal-like and hence anomalous in its basis. As a result one needs not go beyond the harmonic approximation in order to observe anomalous dynamics since anomalous structure leads to it even in the presence of the simplest interactions. Interestingly, when the structure is fractal-like, large conformational changes are encoded within the native state structure. In other words, large conformational changes are implied by the harmonic approximation rather than lead to its breakdown. Anomalous vibrational dynamics in proteins is hence mainly a result of a fractal-like native state structure.

The Fractal Nature of Proteins. Recently there has been an accumulation of evidence supporting the hypothesis that proteins are fractal-like objects (7–16). The fractal nature of proteins manifests itself in the way proteins vibrate and in the manner in which they fill space. Moreover, in a recent study we have shown that proteins are fractals of a very special kind. Proteins obey an equation of state in the parameter space defined by the number of amino acids N and two parameters that characterize the spatial structure: the spectral and fractal dimensions d_s and d_f respectively (12, 13). It appears that when viewed as fractals, proteins exhibit an unexpected universal behavior. The universal scaffold is further equipped with specific chemical ornamentations making proteins the incredible biomachines they are. In light of the fractal view point we introduce a stochastic approach to the vibrational dynamics of proteins. We use diffusion, a method sensitive to the structural features of the protein fold and them alone, in order to probe protein structure. The reason for this is twofold. First, the new method allows us to conduct a large-scale study which revalidates the fractal-like nature of proteins in a manner which is independent of the previous work done in this field. Second, the stochastic approach allows us to provide unique insights and make new predictions regarding single-molecule level protein dynamics.

A Stochastic Approach to Vibrational Dynamics. Studies have shown that it is possible to deduce the innate properties of networks by analyzing the stochastic trajectories of a random walker on them. For fractal networks, it is possible to characterize some basic quantities such as: the probability to be at the origin at time t, the mean square displacement (MSD) from the origin and the mean first passage time (MFPT) between two tagged points, using a small number of parameters (18–21). Interestingly, for ideal fractal networks and within the scalar elasticity model (18) (also known as the Gaussian network model (GNM) when applied for proteins (17)), the same parameters also govern the vibrational properties of the network, when the latter is thought of as a network of masses and springs. Known relations between the vibrational and diffusional properties of networks bridge the gap between network science and protein dynamics at equilibrium. For fractal structures, both the density of vibrational states, $g(\omega)$[†], and the probability of a random walker to be found at the origin at time t, $P_0(t)$, are characterized by the spectral dimension d_s (18, 19):

$$P_0(t) \sim t^{-d_s/2}, \qquad g(\omega) \sim \omega^{d_s-1}. \qquad [1]$$

Author contributions: S.R., R.G., and J.K. designed research; S.R., R.G., and J.K. performed research; S.R. analyzed data; and S.R., R.G., and J.K. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[†]A related quantity is $G(\omega)$, the cumulative density of states defined as $G(\omega) = \int_0^\omega g(\omega')d\omega'$. $G(\omega)$ counts the number of normal modes with frequency less than ω.

[1]To whom correspondence should be addressed. E-mail: shlomireuveni@hotmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002018107/-/DCSupplemental.

For regular three-dimensional structures $d_s = 3$ and one recovers the well known $t^{-3/2}$ behavior for the probability to be at the origin (18), and the ω^2 Debye density of states (22). Recent studies on proteins have already shown that $g(\omega)$ obeys the above power law behavior (7, 8, 11–13). However, in contrast to regular three-dimensional structures, for proteins it is usually found that $d_s < 2$ associating them to a realm of low dimensionality. The vibrational meaning of $d_s < 2$ is an excess of low frequency modes and hence increased flexibility (11–13).

An additional link between vibrational dynamics and diffusion is the relation between the propagation of vibrational energy on a fractal network and the MSD of a random walker. In the overdamped limit, the mean square propagation length of an initially localized structural deformation $\langle \Delta^2(t) \rangle$ and the MSD of a random walker are practically equivalent. Both are characterized by the same parameter, the walk dimension d_w, and the following relations hold*s* (18, 19):

$$\mathrm{MSD}(t) \sim t^{2/d_w}, \qquad \langle \Delta^2(t) \rangle \sim t^{2/d_w}. \qquad [2]$$

Based on general theoretical grounds the above relations have been suggested to hold for proteins as well (23). Molecular dynamics simulations conducted on several proteins have demonstrated that $\langle \Delta^2(t) \rangle$ obeys the above power law behavior (9). In particular, it was found that $d_w > 2$, i.e., the propagation of structural deformations is subdiffusive. These interesting results together with the fact that $P_0(t)$ and $\mathrm{MSD}(t)$ has not yet been studied on proteins, provide motivation for further study.

Diffusion on the Protein Fold

Starting from a coarse-grained model of a protein we study diffusion in order to deduce as much as possible on the vibrational dynamics of proteins. Here we use the GNM (17), a model originally devised for the study of vibrational dynamics. The GNM considers proteins to be elastic networks whose identical nodes correspond to the positions of the alpha-carbons in the native structure and the interactions among nodes are modeled as homogeneous harmonic springs. An interaction between two nodes exists only if the nodes are separated by a distance less than R_c which is known as the interaction cutoff. The cutoff distance is usually taken in the range 6 Å–7 Å, based on the radius of the first coordination shell around residues observed in Protein Data Bank (PDB) structures (24, 25). In this paper we have chosen $R_c = 6$ Å. The only information required to implement the method is the knowledge of the native structure obtained from the PDB (26). The GNM has been widely applied because it yields results in agreement with X-ray spectroscopy and NMR experiments (17, 27). In order to study diffusion, we transform all springs in the GNM into edges creating a three-dimensional network of nodes. On this network, which is a coarse-grained representation of the native structure, we define the following master equation:

$$\frac{d\vec{p}(t)}{dt} = -\frac{\gamma}{\zeta}\Gamma\vec{p}(t). \qquad [3]$$

Here $\vec{p}(t)$ is a vector whose i-th entry is the probability of a random walker to be at the i-th node at time t and the constants γ, ζ are defined in *Methods*. The matrix Γ is the GNM Kirchhoff Matrix. For $i \neq j$, $\Gamma_{ij} = -1$ if i, j are connected by a spring and $\Gamma_{ij} = 0$ otherwise. Γ_{ii} is set to be the coordination number of the i-th node. Eq. **3** describes the stochastic dynamics of a random walker travelling on the protein fold and was defined in the above form in order to preserve the relations between vibrations and diffusion (see *Methods*). From the definition of Γ it follows that the rate at which a random walker jumps out of some node i is proportional to the coordination number of that node and that the random walker performs random jumps to its nearest neighbors without preference. In this scenario, $-\frac{\gamma}{\zeta}\Gamma$ plays the part of a transition rate matrix whose matrix element $-\frac{\gamma}{\zeta}\Gamma_{ij}$ is the rate of transition from state j to state i.

Results

Anomalous Diffusion on the Protein Fold, an Indication of Anomalous Vibrational Dynamics. We solve Eq. **3** for a set of 512 protein structures taken from the PDB. For each protein the equation is solved N times, one time for every amino acid. The initial conditions in each run are probability one to be at the origin (delta function initial conditions) where the origin is taken as the corresponding amino acid for that run. The procedure is discussed further in *Methods*. Averaging over all possible origins, we find that $P_0(t)$ and $\mathrm{MSD}(t)$ are both anomalous. In Fig. 1 we plot $P_0(t)$ for three different proteins. In the limit of very short times $P_0(t) \simeq 1 - \frac{\gamma t}{N\zeta}\sum_{n=1}^{N}\Gamma_{ii}$, in the limit of very long times $P_0(t)$ approaches the stationary distribution (the constant $1/N$) implied by Eq. **3** exponentially fast. We focus on the intermediate time regime for which $P_0(t)$ is found to obey [**1**]. We extract the spectral dimension and find that for most proteins $d_s < 2$ indicating that $P_0(t)$ decays much slower than one would have naively expected. The spectral dimension obtained from this calculation is in good agreement with the one obtained using normal mode analysis (12). Translating this result to a "vibrational" language, we infer (by use of [**1**]) that proteins are characterized by a non-Debye density of vibrational states. The low dynamic dimensionality leads to increased flexibility and was shown to take an important role in bridging two seemingly conflicting properties of native proteins (12, 13). While proteins need to keep their specific native fold structure thermally stable, they must also be able to display flexible motions that allow proper function (5, 28–30). In compact objects which are characterized by small amplitude vibrations and a Debye density of low frequency modes this conflict cannot be bridged. In contrast, the fractal-like structure of proteins was shown central to the coexistence of stability and flexibility (12, 13).

In Fig. 2 we plot $\mathrm{MSD}(t)$ for three different proteins. We focus again on the intermediate time regime for which $\mathrm{MSD}(t)$ is found to obey [**2**]. We extract the walk dimension and find that for pro-

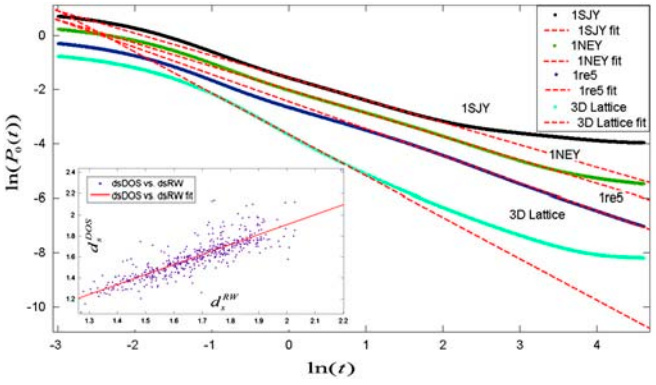


Fig. 1. In proteins, $P_0(t)$—the probability to be at the origin at time t (averaged over all possible origins) obeys a power law relation for intermediate times. We see results for three proteins (PDB codes top to bottom): 1SJ1: $N = 147$, $R_g = 16.2$ Å, 1NEY: $N = 492$, $R_g = 24.37$ Å, and 1re5: $N = 1{,}767$, $R_g = 35.66$ Å. Here N is the number of amino acids and R_g stands for the radius of gyration. On a double log plot the slope corresponds to $-d_s^{\mathrm{RW}}/2$ (RW stand for random walk) and we find: $d_s^{\mathrm{RW}} = 1.64$, $d_s^{\mathrm{RW}} = 1.74$, and $d_s^{\mathrm{RW}} = 2$ correspondingly. For comparison $P_0(t)$ on a finite, 2,197 nodes $R_g = 61.23$ Å, cubic lattice is also shown. Here, a faster decay of $P_0(t)$ in the power law regime ($d_s^{\mathrm{RW}} \simeq 3$) and an earlier crossover to the exponentially decaying regime are clearly visible. For clarity the plots for the three-dimensional Lattice, 1re5, 1NEY, and 1SJ1 were shifted −0.5, 0, +0.5, and +1 along the y-axis correspondingly. The inset shows that the spectral dimension obtained in this way is in good agreement with the spectral dimension obtained via normal modes (denoted DOS) analysis (12): regression line $d_s^{\mathrm{DOS}} = 0.008 + 0.95 \cdot d_s^{\mathrm{RW}}$ with c.c = 0.86.

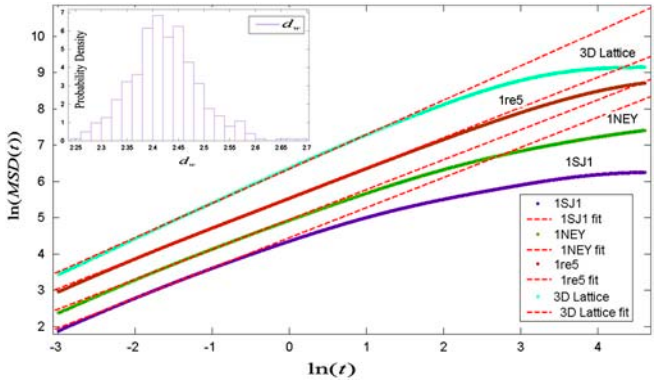


Fig. 2. In proteins, MSD(t)—the mean square displacement of a random walker from the origin (averaged over all possible origins) obeys a power law relation for intermediate times. Here we see results for three proteins (PDB codes bottom to top): 1SJ1 :$N = 147$, $R_g = 16.2$ Å, 1NEY: $N = 492$, $R_g = 24.37$ Å, and 1re5: $N = 1{,}767$, $R_g = 35.66$ Å). On a double log plot the slope corresponds to $2/d_w$ and we find: $d_w = 2.41$, $d_w = 2.42$, and $d_w = 2.40$ correspondingly. The lines were fitted to the region 3.5 Å $\leq \sqrt{\text{MSD}(t)} \leq 0.4R_g$. For comparison MSD($t$) on a finite, 2,197 nodes $R_g = 61.23$ Å, cubic lattice is also shown. Here we find $d_w = 2.11$ indicating that a small part of the anomaly in proteins may be due to finiteness of the structure. For clarity the plots for three-dimensional Lattice, 1re5, 1NEY, and 1SJ1 were shifted +1.5, +1, +0.5, and 0 along the y-axis correspondingly. The inset shows a histogram of the walk dimensions obtained for a set of 512 proteins, d_w is peaked around 2.425 and in all cases $d_w > 2$ which indicates that the motion of a random walker on the protein fold is subdiffusive.

teins $d_w > 2$ indicating subdiffusion. A random walker travelling on the protein fold progress slower than what one would have naively expected. Translating this result to a vibrational language, we infer (by use of [**2**]) a subdiffusive spread of vibrational wave packets. The physical meaning is that on the protein fold the propagation of vibrational energy is slower than normal (9, 31). Moreover, this phenomena cannot be explained by a mere adjustment of the diffusion coefficient since the functional dependence of $\langle \Delta^2(t) \rangle$ on time is no longer linear! Subdiffusion is a characteristic property of random motion on fractal structures and once more the recognition that proteins are characterized by such a structure allows us to reconcile theory and practice.

The path taken in this paper is independent of previous work done in the field and hence allows us to explore the vibrational dynamics of proteins from a new perspective. In particular, we show (see *Methods*) that the average time dependent vibrational MSD can be written as an integral over $P_0(t)$. An experimentally testable prediction follows:

$$\frac{1}{N}\sum_i \langle (\Delta X_i(t) - \Delta X_i(0))^2 \rangle \sim t^{1-d_s/2} \qquad [4]$$

i.e., the time dependent vibrational MSD of amino acids grows subdiffusively with time. Interestingly, molecular dynamics simulations on small proteins have indicated a similar behavior for the MSD of dihedral angles (4).

The First Passage Time Problem on the Protein Fold Leads to Insights on Single-Molecule Level Protein Dynamics. As we have previously discussed proteins are flexible molecules. In quantifiable evidence of this is the experimentally measurable variance $\langle \Delta R_{ij}^2 \rangle \equiv \langle R_{ij}^2 - (R_{ij}^0)^2 \rangle$ in the instantaneous distance R_{ij} between amino acids i and j (32). Here the brackets stand for ensemble average and R_{ij}^0 is the equilibrium distance. This variance is a result of thermal fluctuations and its magnitude is governed by the mechanical properties of the protein. Let us consider the MFPT it takes a random walker, obeying Eq. **3**, to travel from amino acid i to j and vice versa. We define $T(i,j)$ as the average MFPT for the two directions of travel. It has recently been shown that in the GNM there is a simple relation between $T(i,j)$ and $\langle R_{ij}^2 - (R_{ij}^0)^2 \rangle$ (33):

$$\frac{T(i,j)}{N} = \frac{C\gamma}{3k_B T}\langle \Delta R_{ij}^2 \rangle, \qquad [5]$$

where the proportionality constant C is a known numeric factor independent of i and j. On the other hand, a recently obtained expression for the MFPT between two tagged nodes (averaged over all nodes distanced r apart) on a finite fractal structure of N nodes with $d_s < 2$ is given by (20, 21):

$$\frac{T(r)}{N} = A + Br^{\alpha}, \qquad [6]$$

where A and B are structure specific constants and $\alpha = d_w[1 - \frac{d_s}{2}]$. The average MFPT is hence dependent only upon the distance between source and target, the spectral dimension d_s and the walk dimension d_w. A similar conclusion regarding the average variance in the distance between two nodes on a fractal elastic network immediately follows and was also independently derived (23). We have calculated the MFPT between all amino acid pairs for a random walker obeying Eq. **3** on a dataset of 512 proteins (see *Methods*). Our calculations confirmed that for intermediate distances the thermal variance in the instantaneous distance, $\langle \Delta R_{ij}^2 \rangle$, is a power law function of the equilibrium distance R_{ij}^0. Our results are summarized in Fig. 3 and call for experimental validation on real proteins.

The power law dependence of the variance upon distance is significantly different from the one observed for objects characterized by a spectral dimension higher than two (21). In these objects distances between nodes vary within a small range and large conformational changes are hence improbable. Proteins on the other hand, are characterized by a spectral dimension which is lower than two and this leads to a much higher ability to flex. More accurately, when $d_s > 2$, the variance in the instantaneous distance converges to a constant value independent of system size as the equilibrium distance r is increased (21). However when $d_s < 2$, as is the case in the vast majority of proteins, the variance diverges as r is increased (21)! The steep dependence

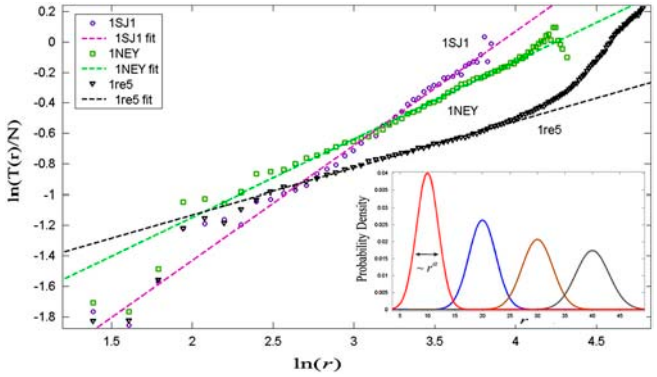


Fig. 3. The MFPT (averaged over all equidistanced pairs) as a function of the distance between the source and target sites. Here the MFPT is normalized by the number of amino acids. We see results for three different proteins (PDB codes: 1SJ1: $N = 147$, $R_g = 16.2$ Å, 1NEY: $N = 492$, $R_g = 24.37$ Å, and 1re5: $N = 1{,}767$, $R_g = 35.66$ Å). On a double log plot the data for intermediate distances is well fitted by straight lines and we conclude that: $\frac{\langle T(r) \rangle}{N} \sim r^{\alpha}$. Here we find that $\alpha = 0.75$, $\alpha = 0.51$, and $\alpha = 0.305$ for PDB codes: 1SJ1, NEY, and 1re5 respectively. The inset is a schematic representation of the vibrational aspect of this phenomena, the instantaneous distance between two amino acids in a protein is distributed around a mean value equal to the native state distance. We predict that on average the width of this distribution grows like a power law of its mean, i.e., there is a power law relation between the native state distance and the magnitude of the fluctuations around this distance. Our prediction and computational validation calls for a complementary experimental validation.

of variance upon distance allows the existence of large conformational changes.

Large-scale conformational changes in proteins, although often associated with the binding of a substrate, have been suggested to be part of the intrinsic dynamics of enzymes in the unbound state (34–36). We now demonstrate how MFPT analysis can be used as a computational tool that may aid in the identification of amino acid pairs involved in such conformational changes. Having determined the functional form of the average MFPT, $T(r)$, we focus on the importance of deviations from the average trend. A rearranged version of Eq. **5**:

$$k_B T = \frac{NC\gamma}{3T(i,j)}\langle \Delta R_{ij}^2 \rangle \equiv \frac{1}{2} k_{ij}^{\text{eff}} \langle \Delta R_{ij}^2 \rangle, \qquad [7]$$

immediately lends itself to an effective inverse spring constant interpretation of $T(i,j)$. The larger the $T(i,j)$ the softer the effective spring connecting residues i and j. These residues are hence more easily brought together (or pushed apart) by the thermal noise. A case where the equilibrium distance between i and j is r and $T(i,j)$ is considerably larger than the $T(r)$ (the average MFPT for that distance) may be indicative of an amino acid pair involved in large-scale conformational changes. In Fig. 4 we demonstrate how this idea can be used to predict which residues are involved in large conformational transitions. At this point we were unable to validate the theoretical relation between α, d_s, and d_w most probably due to finite size effects.

Discussion

In this paper we introduced a stochastic approach to the vibrational dynamics of proteins. We showed that observed anomalies in the vibrational dynamics of proteins are a direct result of a fractal-like structure. Studying diffusion on the protein fold we have isolated the contribution of anomalous structure to the existence of: non-Debye density of vibrational states; anomalous propagation of vibrational deformations; flexibility; and large conformational changes. Tested on over 500 PDB structures our observations appear universal. Our findings highlight the fractal-like structure of proteins positioning it as the central player involved in determining vibrational dynamics.

We have also provided unique experimentally testable predictions regarding the subdiffusive behavior of the time dependent vibrational mean square displacement of amino acids and the thermal variance characterizing the instantaneous interresidual distance. In the past two decades it has been recognized that tracing the MSD of a tagged point on a manifold, be it a polymer or a membrane (37, 38), can provide a direct probe of its energetic and geometric properties. A measurement of the time dependent vibrational MSD of an amino acid may serve as a direct probe for the geometrical features and underlining elastic energy of the protein fold. The determination of interresidual distance distributions by ensemble and single-molecule FRET experiments enabled the exploration of domain motions and conformational fluctuations of enzyme molecules (32). Our findings fall in line with the observation of wide distance distributions (32). Nevertheless, a verification of the universal quantitative features of these distributions, as elucidated above, is only possible by conducting a large-scale experimental study.

We have offered the use of mean first passage time analysis as a practical, computational tool that may aid in the identification of amino acid pairs involved in large conformational changes. Recently, emphasis has been given to geometry controlled reaction kinetics (39). In particular, the role of the first passage time and its moments in determining reaction kinetics on fractal geometries has been elucidated. Our findings regarding the fractal nature of proteins and the resulting MFPT are thus directly linked to diffusion controlled reaction kinetics (40) on the protein fold.

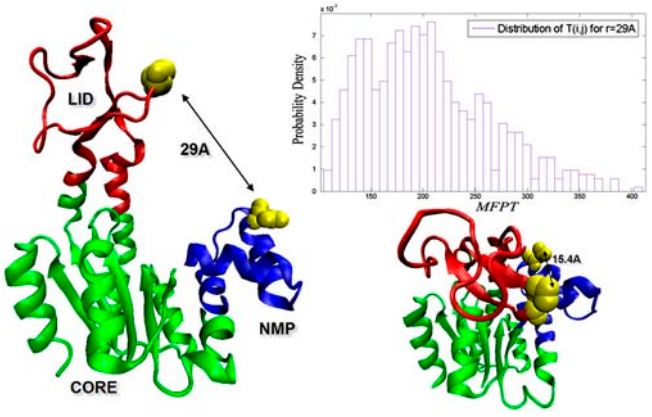


Fig. 4. The enzyme Adenylate kinase in the open (*Left*, PDB code 4AKE) and closed (*Bottom Right*, PDB code 1AKE). In the open state residues 53 and 127, located at the NMP (*blue*) and LID (*red*) domains respectively, are distanced 29 Å apart. For the enzyme in the open state $T(r = 29\text{ Å}) = 204.54$ with a standard deviation of $\sigma = 59.54$. In comparison, $T(53,127) = 337$, a value considerably larger than the mean value in its distance group. This fact is also visible from the histogram of the $T(i,j)$'s in this distance group (*Top Right*). The large value of $T(53,127)$ indicates that this enzyme readily samples conformations in which the distance between residues 53 and 127 is considerably smaller than their native state distance. This fact implies the possible existence of a conformational change encoded within the native structure. Indeed, upon the binding of ATP and AMP the enzyme assumes the closed state in which residues 53 and 127 are only 15.45 Å apart. In reference to the appropriate distance group, abnormally high values of $T(i,j)$ may hence be indicative of amino acid pairs taking part in large conformational transitions. On the other hand, abnormally low values of $T(i,j)$ indicate cooperative motion of an amino acid pair.

Methods

Connecting Diffusion with Vibrations in the GNM. The GNM is an isotropic model, all three spacial directions are equivalent. Focusing on one of the axis, the equations of motion for a protein in the GNM are given by: $M\frac{d^2\Delta\vec{X}}{dt^2} = -\gamma\Gamma\Delta\vec{X}$. Here, $\Delta\vec{X}$ is the $N\times 1$ vector of displacements from equilibrium positions, γ and M are the GNM spring constant and mass respectively, and Γ is the network Kirchhoff matrix. When damping and thermal noise, in the form of $-\zeta\frac{d\vec{X}}{dt}$ and $\zeta\vec{g}(t)$, are introduced to the right hand side of the equation of motion and the high damping limit is considered we get: $\frac{d\Delta\vec{X}}{dt} = -\frac{\gamma}{\zeta}\Gamma\Delta\vec{X} + \vec{g}(t)$. Taking the ensemble average of both sides we conclude that the equations of motion for the average displacements are given by:

$$\frac{d\langle\Delta\vec{X}\rangle}{dt} = -\frac{\gamma}{\zeta}\Gamma\langle\Delta\vec{X}\rangle. \qquad [8]$$

Note that Eqs. **3**, **8** are equivalent and it is now clear why we have chosen the jump rate in Eq. **3** to be proportional to γ/ζ. [**2**] in the paper is an immediate conclusion from the combination of Eqs. **3**, **8** with the fact that on fractal structures $\text{MSD}(t) \sim t^{2/d_w}$ (18).

Solving the Master Equation. In order to compute $P_0(t)$ and $\text{MSD}(t)$ for PDB structures we have solved Eq. **3** numerically using Matlab, taking $\gamma/\zeta = 1$ for convenience. Note that this choice will not change the scaling laws of any of the quantities mentioned above. The solution of Eq. **3** is given by: $\vec{p}(t) = \exp[-t\Gamma]\vec{p}(t=0)$ and we have calculated the matrix exponent $\exp[-t\Gamma]$ for different times. We denote by $\vec{p}^{(i)}(t=0)$ the delta function initial conditions: $p_j(t=0) = \delta_{ij}$ and by $\vec{p}^{(i)}(t)$ the corresponding solution. We have calculated $\vec{p}^{(i)}(t)$ for different times and repeated this process N times (the number of amino acids) going over all possible origins $1 \le i \le N$ (native positions of the C-alpha atoms) for every protein in our dataset. After obtaining the probability distribution $\vec{p}^{(i)}(t)$, the mean square displacement from the origin was calculated using its definition: $\text{MSD}^{(i)}(t) = \sum_{j=1}^{N} p_j^{(i)}(t)(R_{ij}^0)^2$, where R_{ij}^0 is the distance of site j from site i (the origin). The probability to be at the origin $P_0(t)$ and the mean square displacement $\text{MSD}(t)$ were defined as the corresponding averages: $P_0(t) \equiv \frac{1}{N}\sum_{i=1}^{N} p^{(i)}(t)$, $\text{MSD}(t) \equiv \frac{1}{N}\sum_{i=1}^{N}\text{MSD}^{(i)}(t)$.

Derivation of Equation 4. In the *SI Text* provided online we show that the relation between $P_0(t)$ and the average vibrational MSD is given by:

$$\frac{1}{N}\sum_i \langle(\Delta X_i(t)-\Delta X_i(0))^2\rangle = \frac{2k_BT}{\zeta}\int_0^t P_0(t')dt'. \qquad [9]$$

[**4**] is then obtained by use of [**1**].

Calculating the MFPT. In order to calculate the MFPT it takes a random walker to get from site i to site j on the protein fold we consider a random walker starting its voyage at site i, i.e.,: (delta function initial conditions $p_j(0)=\delta_{ij}$), and set $\gamma/\zeta=1$ for convenience. The mean time it takes a random walker, obeying Eq. **3** in the paper, to leave site i is given by: $1/\Gamma_{ii}$ and immediately after this time the random walker has probability $-\Gamma_{ik}/\Gamma_{ii}$ to be found at site k. Because Γ_{ii} is the coordination number of site i, the random walker has an equal probability of $1/\Gamma_{ii}$ to be found at any one of its nearest neighbors and zero probability to be found anywhere else. Denoting by $t(j,i)$ the MFPT it takes a random walker to get from site i to site j (note that in the paper we have defined: $T(i,j)=[t(i,j)+t(j,i)]/2$) we write it in the following recursive form: $t(j,i)=1/\Gamma_{ii}+\sum_k t(j,k)/\Gamma_{ii}$, where we sum only over nearest neighbors and define $t(j,j)\equiv 0$ for every site j. Writing this equation for every source site i keeping the target site j fixed results in a set of N linear equations. Solving them we find $t(j,i)$ for every $1\le i\le N$, repeating the processes for every j we find $t(j,i)$ for every i and j. For $r\in\{1$ Å,2 Å,3 Å…$\}$, $T(r)$ was defined to be the average MFPT between all pairs whose equilibrium distance R_{ij}^0 falls in the range $r-\frac{1}{2}$ Å $\le R_{ij}^0\le r+\frac{1}{2}$ Å.

Remark on the Numeric Constant "C" in Equation 5. In ref. 33 the random walk was performed in discrete time and the authors defined $T(i,j)$ as the sum of the MFPT in both directions of travel. As a result the numeric constant "C" in [**5**] was calculated to be equal to the average coordination number of a node in the network. For consistency with the definition used in this paper we note that for a random walker obeying Eq. **3** (taking $\gamma/\zeta=1$) the numeric constant "C" is simply given by: $C=\frac{1}{2}$.

Calculated Exponents. Calculated exponents (d_s, d_w, α) for Dataset S1 are given in a separate Excel file to be found online.

ACKNOWLEDGMENTS. S.R. acknowledges support from the Converging Technologies program of the Israeli Council for higher education. R.G. and J.K. acknowledge support from the Israel Science Foundation and the Deutsch-Israelische Projektkooperation (DIP). R.G. is a member of the Ilse Katz Center for Meso and Nanoscale Science and Technology and the Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering.

1. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science 254* 5038:1598–1603.
2. Sabelko J, Ervin J, Gruebele M (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA* 96:6031–6036.
3. Bruji J, Hermans RIZ, Walther KA, Fernandez JM (2006) Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nature Physics* 2:282–286.
4. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA (2008) How main-chains of proteins explore the free-energy landscape in native states. *Proc Natl Acad Sci USA* 105:19708–19713.
5. Henzler-Wildman K, Keren D (2007) Dynamic personalities of proteins. *Nature* 450:964–972.
6. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Observation of a power-law memory kernel for fluctuations within a single protein molecule. *Phys Rev Lett* 94:198302.
7. Stapleton HJ, Allen JP, Flynn CP, Stinson DG, Kurtz SR (1980) Fractal form of proteins. *Phys Rev Lett* 45:1456–1459.
8. Lushnikov SG, Svanidze AV, Sashin IL (2005) Vibrational density of states of hen egg white lysozyme. *JETP Letters* 82:30–33.
9. Yu X, Leitner DM (2003) Anomalous diffusion of vibrational energy in proteins. *Journal of Chemical Physics* 119:12673.
10. Banerji A, Ghosh I (2009) Revisiting the myths of protein interior: studying proteins with mass-fractal hydrophobicity-fractal and polarizability-fractal dimensions. *PLoS ONE* 4:e7361 doi:10.1371/journal.pone.0007361.
11. Burioni R, Cassi D, Cecconi F, Vulpiani A (2004) Topological thermal instability and length of proteins. *Proteins* 55:529–535.
12. Reuveni S, Granek R, Klafter J (2008) Proteins: coexistence of stability and flexibility. *Phys Rev Lett* 100:208101.
13. de Leeuw M, Reuveni S, Klafter J, Granek R (2009) Coexistence of flexibility and stability of proteins: an equation of state. *PLoS ONE* 4:e7296 doi:10.1371/journal.pone.0007296.
14. Chowdary PD, Gruebele M (2009) Molecules: what kind of a bag of atoms? *J Phys Chem A* 113:13139–13143.
15. Enright MB, Leitner DM (2005) Mass fractal dimension and the compactness of proteins. *Phys Rev E* 71:011912.
16. Elber R, Karplus M (1986) Low-frequency modes in proteins: use of the effective-medium approximation to interpret the fractal dimension observed in electron-spin relaxation measurements. *Phys Rev Lett* 56:394–397.
17. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 2:173–181.
18. Nakayama T, Yakubo K *Fractal concepts in condensed matter physics* (Springer-Verlag, Berlin), pp 59–88, 186–191.
19. Nakayama T, Kousuke Y, Orbach RL (1994) Dynamical properties of fractal networks: scaling, numerical simulations, and physical realizations. *Rev Mod Phys* 66:381–443.
20. Condamin S, Bénichou O, Tejedor V, Voituriez R, Klafter J (2007) First-passage times in complex scale-invariant media. *Nature* 450:77–80.
21. Reuveni S, Granek R, Klafter J (2010) Vibrational shortcut to the mean-first-passage-time problem. *Phys Rev E* 85:040103(R).
22. Ashcroft NW, Mermin ND (1976) *Solid State Physics* (Harcourt, Orlando), pp 464–466.
23. Granek R, Klafter J (2005) Fractons in proteins: can they lead to anomalously decaying time autocorrelations? *Phys Rev Lett* 95:098106.
24. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18:534–552.
25. Bahar I, Jernigan RL (1997) Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separations. *Journal of Molecular Biology* 266:195–214.
26. Bernstein FC, et al. (1977) The protein data bank: a computer-based archival file for macro-molecular structure. *J Mol Biol* 112:535–542.
27. Yang LW, et al. (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure* 15:741–749.
28. Joseph D, Petsko GA, Karplus M (1990) Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science* 249:1425.
29. Steinbach PJ, et al. (1991) Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry* 30:3988.
30. Rasmussen BF, Stock AM, Ringe D, Petsko GA (1992) Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* 357:423.
31. Leitner DM (2008) Energy flow in proteins. *Annu Rev Phys Chem* 59:233–259.
32. Haas E (2005) The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. *Chem Phys Chem* 6:858–870.
33. Chennubhotla C, Bahar I (2007) Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput Biol* 3:e172 doi:10.1371/journal.pcbi.0030172.
34. Haran G, Haas E, Szpikowska BK, Mas MT (1992) Domain motions in phosphoglycerate kinase: determination of interdomain distance distributions by site-specific labeling and time-resolved fluorescence energy transfer. *Proc Natl Acad Sci USA* 89:11764–11768.
35. Arora K, Brooks CL, III (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci USA* 104:18496–18501.
36. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic enzyme motions in the unbound state and relation to allosteric regulation. *Curr Opin Struc Biol* 17:633–640.
37. Granek R (1997) From semi-flexible polymers to membranes: anomalous diffusion and reptation. *J Phys* 7:1761–1788.
38. Caspi A, Elbaum M, Granek R, Lachish A, Zbaida D (1998) Semiflexible polymer network: a view from inside. *Phys Rev Lett* 80:1106–1109.
39. Bénichou O, Chevalier C, Klafter J, Meyer B, Voituriez R (2010) Geometry-controlled kinetics. *Nature Chemistry* DOI: 10.1038/NCHEM.622.
40. Shinobu A, Agmon N (1997) Mapping proton wires in proteins: carbonic anhydrase and GFP chromophore biosynthesis. *Phys J Chem A* 113:7253–7266.

Supporting Information

Reuveni et al. 10.1073/pnas.1002018107

SI Text

A Derivation of Equation 4 in the Paper. ***Decoupling the equations of motion.*** We start from the Langevin equation for the GNM in the high damping limit:

$$\frac{d\Delta\vec{X}}{dt} = -\frac{\gamma}{\zeta}\Gamma\Delta\vec{X} + \vec{g}(t), \qquad [S1]$$

here the random forces are characterized by (S1):

$$\langle\vec{g}(t)\rangle = \vec{0}, \qquad \langle g_i(t)g_j(t')\rangle = \frac{2k_BT}{\zeta}\delta_{ij}\delta(t-t'). \qquad [S2]$$

A brief look at this equation reveals that the variables are coupled. The dynamics of the i-th amino acid, described by the deviation from equilibrium ΔX_i, depends not only on ΔX_i itself but also on other nodes $\{\Delta X_j\}$. Although the above description is very natural, since the coordinates used are the actual deviations from equilibrium, it leads to a rather complicated set of equations. It is sometimes beneficial to describe the system differently using a special set of coordinates called normal coordinates. Describing the system with normal coordinates leads to N uncoupled equations of motion.

The matrix Γ is real and symmetric* by definition. One of the basic theorems concerning such matrices is the finite-dimensional spectral theorem, which says that any symmetric matrix whose entries are real can be diagonalized by a real orthogonal matrix†. More explicitly: to every symmetric real matrix Γ there exists a real orthogonal matrix A such that $D = A^{-1}\Gamma A = A^T\Gamma A$ is a diagonal matrix. Let A be the real orthogonal matrix (whose columns $\{\vec{A}_j\}$ are the eigenvectors of Γ) that diagonalizes the matrix Γ. We define a new set of coordinates $\{\Delta U_i\}$ using the old set of coordinates $\{\Delta X_i\}$ by the orthogonal transformation:

$$\Delta\vec{X} = A\Delta\vec{U}. \qquad [S3]$$

Combining Eqs. **S1**, **S3** we get:

$$A\frac{d\Delta\vec{U}}{dt} = \frac{dA\Delta\vec{U}}{dt} = \frac{d\Delta\vec{X}}{dt} = -\frac{\gamma}{\zeta}\Gamma\Delta\vec{X} + \vec{g}(t) = -\frac{\gamma}{\zeta}\Gamma A\Delta\vec{U} + \vec{g}(t), \qquad [S4]$$

multiplying both sides by A^{-1} we obtain the equations of motion for the normal coordinates $\{\Delta U_i\}$:

$$\frac{d\Delta\vec{U}}{dt} = -\frac{\gamma}{\zeta}D\Delta\vec{U} + \tilde{g}(t). \qquad [S5]$$

Here $D = A^{-1}\Gamma A$ is a diagonal matrix whose entries are the eigenvalues $\{\omega_1^2,\omega_2^2,\ldots,\omega_N^2\}$ of the matrix Γ. We note that $\omega_1^2 = 0$ (S2) (a normal mode that corresponds to translatory motion of the protein) and that $\tilde{g}(t) \equiv A^{-1}\vec{g}(t)$. From this definition it follows that new random forces are still characterized by:

$$\langle\tilde{g}(t)\rangle = \vec{0}, \qquad \langle\tilde{g}_i(t)\tilde{g}_j(t')\rangle = \frac{2k_BT}{\zeta}\delta_{ij}\delta(t-t'), \qquad [S6]$$

where we have used: $\langle\tilde{g}(t)\rangle = \langle A^{-1}\vec{g}(t)\rangle = A^{-1}\langle\vec{g}(t)\rangle = \vec{0}$ and $\langle\tilde{g}(t)\tilde{g}(t')^T\rangle = \langle A^{-1}\vec{g}(t)\vec{g}(t')^TA\rangle = A^{-1}\langle\vec{g}(t)\vec{g}(t')^T\rangle A = \frac{2k_BT}{\zeta}\delta(t-t')$ $A^{-1}IA = \frac{2k_BT}{\zeta}\delta(t-t')I$. Rewriting Eq. **S5** in a non matrix form, we find that:

$$\frac{d\Delta U_i}{dt} = -\frac{\gamma\omega_i^2}{\zeta}\Delta U_i + \tilde{g}_i(t), \qquad [S7]$$

ΔU_i thus obey the Langevin equation of a simple harmonic oscillator whose spring constant is given by: $k_i = \gamma\omega_i^2$.

The time dependent mean square displacement in the GNM. We will now calculate the mean square displacement of the i-th amino acid:

$$\langle(\Delta X_i(t) - \Delta X_i(0))^2\rangle = \langle(\Delta X_i(t))^2\rangle + \langle(\Delta X_i(0))^2\rangle - 2\langle\Delta X_i(t)\Delta X_i(0)\rangle. \qquad [S8]$$

We start by noting that the right hand side of this equation may be written as:

$$\langle(\vec{A}_i^T\cdot\Delta\vec{U}(t))^2\rangle + \langle(\vec{A}_i^T\cdot\Delta\vec{U}(0))^2\rangle - 2\langle(\vec{A}_i^T\cdot\Delta\vec{U}(t))(\vec{A}_i^T\cdot\Delta\vec{U}(0))\rangle, \qquad [S9]$$

where $\vec{A}_i^T$ is the i-th row (column) of A (A^T). In order to proceed we recall that for $i \neq j$: $\langle\Delta U_i(t)\Delta U_j(t')\rangle = \langle\Delta U_i(t)\rangle\langle\Delta U_j(t')\rangle$ due to the independence of the normal coordinates. Because it is also true that (S1) $\langle\Delta\vec{U}(0)\rangle = \langle\Delta\vec{U}(t)\rangle = \vec{0}$ the correlation function $\langle\Delta U_i(t)\Delta U_j(t')\rangle$ vanishes altogether for $i \neq j$. Defining the origin as the position of the center of mass at time t (i.e., $\Delta U_1(0) \equiv 0$) and recalling that (S1):

$$\begin{cases}\langle\Delta U_i(t)\Delta U_i(0)\rangle = \frac{k_BT}{k_i}e^{-(k/\zeta)t} & i > 1\\ \langle\Delta U_1(t)^2\rangle = \frac{2k_BT}{\zeta}t & \\ \langle\Delta U_i(t)^2\rangle = \langle\Delta U_i(0)^2\rangle = \frac{k_BT}{k_i} & i > 1\end{cases}. \qquad [S10]$$

It is straight forward to show that for $t > 0$ Eq. **S8** can be brought to the following form:

$$\langle(\Delta X_i(t) - \Delta X_i(0))^2\rangle = \frac{2k_BTA_{i1}^2}{\zeta}t + \sum_{j=2}^{N}\frac{2k_BTA_{ij}^2}{\gamma\omega_j^2}[1 - e^{-\frac{\gamma\omega_j^2}{\zeta}t}]. \qquad [S11]$$

We note that the first term is due to the diffusion of the protein's center of mass and that the rest of the terms are due to vibrations. The average mean square displacement is given by:

$$\frac{1}{N}\sum_{i=1}^{N}\langle(\Delta X_i(t) - \Delta X_i(0))^2\rangle = \frac{2k_BT}{N\zeta}t + \sum_{j=2}^{N}\frac{2k_BT}{N\gamma\omega_j^2}[1 - e^{-\frac{\gamma\omega_j^2}{\zeta}t}], \qquad [S12]$$

where we have used the fact that the columns of the matrix A form an orthonormal basis and hence $\sum_{i=1}^{N}A_{ij}^2 = 1$ for every j.

The probability to be at the origin. In order to relate the probability to be at the origin with the time dependent mean square displa-

*In linear algebra, a symmetric matrix is a square matrix Γ, that is equal to its transpose $\Gamma = \Gamma^T$. The entries of a symmetric matrix are symmetric with respect to the main diagonal (top left to bottom right), so if the entries are written as a_{ij}, then $a_{ij} = a_{ji}$.

†In matrix theory, a real orthogonal matrix is a square matrix A whose transpose is its inverse: $A^TA = AA^T = I$. A real square matrix is orthogonal if and only if its columns form an orthonormal basis of the Euclidean space R^N with the ordinary Euclidean dot product, which is the case if and only if its rows form an orthonormal basis of R^N.

cement we go back to Eq. **3** in the paper:

$$\frac{d\vec{p}(t)}{dt} = -\frac{\gamma}{\zeta}\Gamma\vec{p}(t). \qquad \textbf{[S13]}$$

Because the eigenvectors (of the matrix Γ) $\{\vec{A}_j\}$ form an orthonormal base of R^N any initial condition, $\vec{p}(0)$, can be expressed as:

$$\vec{p}(0) = \sum_{j=1}^{N} a_j\vec{A}_j. \qquad \textbf{[S14]}$$

Here, the coefficient a_i is found by taking the dot product of $\vec{A}_i$ with both sides of the equation and using orthonormality ($\vec{A}_i \cdot \vec{A}_j = \delta_{ij}$):

$$a_i = \vec{A}_i \cdot \vec{p}(0). \qquad \textbf{[S15]}$$

The solution of Eq. **S13** may now be written in terms of the eigenvectors $\{\vec{A}_j\}$:

$$\vec{p}(t) = e^{-\frac{\gamma}{\zeta}\Gamma}\vec{p}(0) = \sum_{j=1}^{N} a_j e^{-\frac{\gamma\omega_j^2 t}{\zeta}}\vec{A}_j. \qquad \textbf{[S16]}$$

In particular if the random walker starts from the i-th node with probability 1 then $\vec{p}(0)$ is a column vector whose entries are all zero except the i-th entry which equals 1 (delta function initial conditions), in that case we get:

$$p_j(0) = \delta_{ij}, \qquad \vec{p}(t) = e^{-\frac{\gamma}{\zeta}\Gamma}\vec{p}(0) = \sum_{j=1}^{N} A_{ij} e^{-\frac{\gamma\omega_j^2 t}{\zeta}}\vec{A}_j. \qquad \textbf{[S17]}$$

It follows that the probability of a random walker to be at the origin (i-th node) at time t is given by:

$$p_i(t) = \sum_{j=1}^{N} A_{ij}^2 e^{-\frac{\gamma\omega_j^2 t}{\zeta}}. \qquad \textbf{[S18]}$$

Deriving equation 4. Integrating over Eq. **S18** we find that for $t > 0$:

$$\int_0^t p_i(t')dt' = A_{ij}^2 t + \sum_{j=2}^{N} \frac{\zeta A_{ij}^2}{\gamma\omega_j^2}\left[1 - e^{-\frac{\gamma\omega_j^2 t}{\zeta}}\right]. \qquad \textbf{[S19]}$$

Eq. **S19** has a striking similarity to Eq. **S11**. Indeed, we have found that the time dependent mean square displacement of the i-th amino acid may be written as an integral over the probability to be at the origin:

$$\langle(\Delta X_i(t) - \Delta X_i(0))^2\rangle = \frac{2k_B T}{\zeta}\int_0^t p_i(t')dt'. \qquad \textbf{[S20]}$$

Summing over all amino acids and dividing by N we obtain a similar expression for the average mean square displacement:

$$\frac{1}{N}\sum_{i=1}^{N}\langle(\Delta X_i(t) - \Delta X_i(0))^2\rangle = \frac{2k_B T}{\zeta}\int_0^t P_0(t')dt'. \qquad \textbf{[S21]}$$

The scaling law for the vibrational part of the average mean square displacement (Eq. **4** in the paper) is obtained by substituting Eq. **1** in the paper:

$$P_0(t) \sim t^{-d_s/2}, \qquad \textbf{[S22]}$$

into Eq. **S21**. Note that the average mean square displacement has yet another trivial contribution associated with diffusion of the protein as a whole. The diffusive term is linear with time as was mentioned after Eq. **S11**.

Other Supporting Information Files

Dataset S1 (XLS)

1. Doi M, Edwards SF (1988) *The theory of polymer dynamics* (Oxford University Press, USA), 46–55.

2. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 2:173–181.

Dynamic Structure Factor of Vibrating Fractals

Shlomi Reuveni,[1,2] Joseph Klafter,[1] and Rony Granek[3,*]

[1]School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel

[2]School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

[3]The Stella and Avram Goren-Goldstein Department of Biotechnology Engineering, Ben-Gurion University of The Negev, Beer Sheva 84105, Israel

(Received 18 September 2011; published 7 February 2012)

Motivated by novel experimental work and the lack of an adequate theory, we study the dynamic structure factor $S(k, t)$ of large vibrating fractal networks at large wave numbers k. We show that the decay of $S(k, t)$ is dominated by the spatially averaged mean square displacement of a network node, which evolves subdiffusively in time, $\langle(\vec{u}_i(t) - \vec{u}_i(0))^2\rangle \sim t^{\nu}$, where ν depends on the spectral dimension d_s and fractal dimension d_f. As a result, $S(k, t)$ decays as a stretched exponential $S(k, t) \approx S(k)e^{-(\Gamma_k t)^{\nu}}$ with $\Gamma_k \sim k^{2/\nu}$. Applications to a variety of fractal-like systems are elucidated.

DOI: 10.1103/PhysRevLett.108.068101 PACS numbers: 05.45.Df, 05.40.−a, 47.53.+n

Naturally occurring fractals are ubiquitous [1]. Fractal models have been used to describe the dynamics of low temperature glasses and porous materials [2], proteins [3–8], sol-gel branched polymer clusters [9], and colloidal aggregates [10]. Scattering experiments, in which one is able to simultaneously probe correlations in space and time, allow the characterization of fractal structures. A key player in these experiments is the structure factor (SF) [11,12]. While the static SF of fractals is well understood [2,11,12], dynamic structure factor (DSF) calculations are limited. In the context of solid fractals, the DSF has been extensively analyzed on the ''single phonon'' level [2], and in the absence of any source of friction. This provides a good description for the inelastic (Brillouin) scattering from solid fractals, but is not adequate for the quasielastic scattering from low dimensional fractals in solutions that have large fluctuations and friction dominated dynamics, such as branched polymers and colloidal aggregates [9,10].

In this Letter we calculate the DSF $S(k, t)$ of vibrating fractal structures. A striking example of a biological fractal in solution is the spatial organization of chromatin in the nucleus. Recent experiments [13] stand in line with long-standing theoretical predictions [14] suggesting a fractal (crumpled) globule structure. And yet, other fractal chromatin structures were also claimed consistent with experimental results [15]. Combined with the theory presented herein, DSF studies may aid in resolving this conflict. Interestingly, the DSF has been measured in a different biological system in which neutron spin-echo studies were performed on horse heart myoglobin and bovine hemoglobin in solutions. In the large wave number k regime corresponding to $kR_g \gg 1$, where R_g is the gyration radius, and at low concentrations and times shorter than 1 ns, the DSF decays as a stretched exponential [16]. As we demonstrate here, this decay is distinctive of fractal structures, and the fractal-like nature of proteins [4–6,8] makes them a natural case study for our theory. A stretched exponential decay of the DSF is also observed in dynamic light scattering experiments from suspensions of soft colloids that form glasses at large volume fractions. In these systems a universal stretching exponent that is independent of the volume fraction is found [17]. These findings can be explained by the theory advanced here assuming similarity of the structure (or, more precisely, of the force constant network) between these glasses and 3D percolation networks, as done for solid glasses [2].

Fractals are characterized by a few broken dimensions [18]: (i) the mass fractal dimension d_f, that governs the scaling $M(r) \sim r^{d_f}$ of the mass $M(r)$ enclosed in concentric spheres of radius r, (ii) the topological dimension d_l that governs the scaling $M(l) \sim l^{d_l}$ of the mass $M(l)$ enclosed in concentric ''spheres'' of radius l in the topological (or ''manifold'' or ''chemical'') space, and (iii) the spectral dimension d_s that governs the scaling $g(\omega) \sim \omega^{d_s-1}$ of the vibrational density of states $g(\omega)$ with frequency ω [2]. The fractal dimension is experimentally measurable as it governs the power law behavior of the static SF. While known to be related, a general theory explaining the decay of the DSF in terms of d_f and d_s is still lacking and elucidation of d_s based on the DSF is usually not possible.

The main result of this Letter can be simply stated as follows. Assume a large bead-spring fractal network with $d_s < 2$ and arbitrary d_f, and consider a scattering experiment at large wave numbers k such that both $kR_g \gg 1$ and $k\bar{u} \gg 1$, where $\bar{u} \equiv \sqrt{\langle u^2\rangle}$ is the root mean square bead displacement. The latter obeys the generalized Landau-Peierls instability [5,6], $\bar{u} \sim N^{1/d_s-1/2}$, where N is the number of beads. In this limit, and within a wide window of time, we find that $S(k, t)$ decays as a stretched exponential $S(k, t) \simeq S(k)e^{-(\Gamma_k t)^{\nu}}$, where the relaxation rate anomalously depends on k, $\Gamma_k \sim k^{2/\nu}$. The stretched exponential relaxation is a consequence of the anomalous diffusion of a

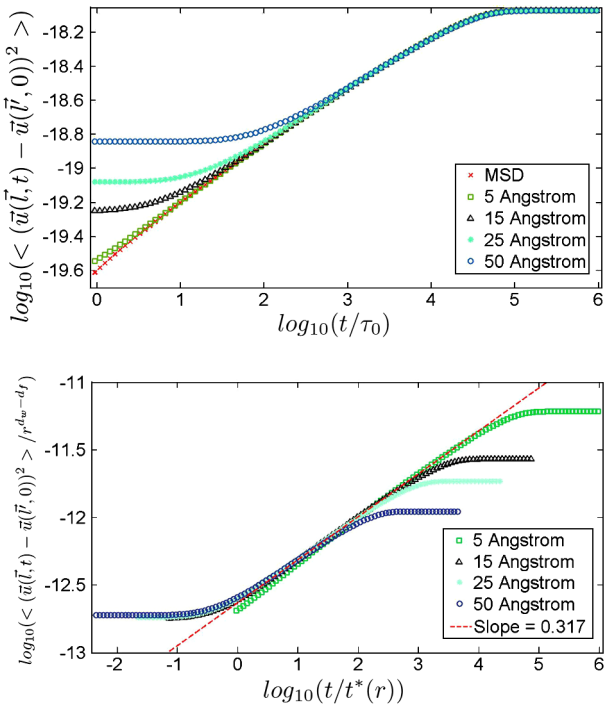


FIG. 1 (color online). (a) The pair correlation function $\langle(\vec{u}(\vec{l},t)-\vec{u}(\vec{l}',0))^2\rangle$ is evaluated numerically, using the Rouse-type model, for bead pairs located on a vibrating Sierpinski gasket with 6561 nodes [25]. For a *fixed* value of the interbead distance r, $\langle(\vec{u}(\vec{l},t)-\vec{u}(\vec{l}',0))^2\rangle$ is calculated for all pairs distanced $r\pm\frac{1}{2}$ Å apart where $r=5, 15, 25, 50$ Å. For every point in time, the correlation functions, in each distance group, are averaged over all pairs in that group. (b) Following the predicted scaling behavior stated in Eq. (7), we normalize the correlation functions from (a) by $r^{d_w-d_f}$ and the time by $t^*(r)\sim r^{d_w}$. Data collapse to a single master curve is observed for $\tau_0 \ll t \ll \tau_N$. The slope ν in the subdiffusive time regime is found to be 0.317, in excellent agreement with the theoretical value of $\nu = 1 - d_s/2 \simeq 0.317$.

network bead, with a mean square displacement (MSD) evolving as $\sim t^{\nu}$. The exponent ν depends on the fractal and spectral dimensions, $\nu = 1 - d_s/2$ in a Rouse model where the friction is local [3,6,19,20], $\nu = (2-d_s)/(2-d_s+d_s/d_f)$ in a Zimm model where the friction is long range [7], and $\nu = 2 - d_s$ for vanishing friction [3]. The latter situation applies for solid fractals, yielding a result which goes much beyond the "single phonon or fracton" approach used previously and valid only for $k\bar{u} \ll 1$ [2]. Our result allows for an experimental evaluation of d_s.

We repeat briefly the definitions and assumptions of the scalar elasticity model [2]. The ground configurational state of the fractal is described by the set of coordinates $\vec{R}_{\rm eq}(\vec{l})$, where $\vec{l}$ is the coordinate of a bead in topological space, and deviations from the ground state are denoted by the displacements $\vec{u}(\vec{l}) = \vec{R}(\vec{l}) - \vec{R}_{\rm eq}(\vec{l})$. The scalar elasticity "bead-spring" Hamiltonian is [2]

$$H[\{\vec{u}(\vec{l})\}] = \frac{1}{2} m\omega_0^2 \sum_{\langle\vec{l}\vec{l}'\rangle} (\vec{u}(\vec{l}) - \vec{u}(\vec{l}'))^2, \quad (1)$$

where $\langle\vec{l}\vec{l}'\rangle$ stands for pairs connected by springs, ω_0 is the spring self-frequency, and m is the bead mass ($m\omega_0^2$ is the spring constant). The eigenstates (normal modes) of the Hamiltonian (1) form an orthonormal set and, on a fractal, are strongly localized in space, bearing the name "fractons" [2,21].

In order to derive the DSF of vibrating fractals, we discuss first the relevant displacement pair correlation function $\langle(\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0))^2\rangle$, where it is understood that spatial averaging has been performed, thus making this correlation function depend only on the relative separation $|\vec{\ell}-\vec{\ell}'|$ in topological space.

In Refs. [3,6,7] we derived the normal mode space Langevin equations for fractals in the high damping and vanishing damping limits. In the high damping limit, which is our main focus here due to its relevance to fractal dynamics in solutions, two models were considered [11,20]: (i) a Rouse-type model in which the hydrodynamic friction is local [3] and (ii) a Zimm-type model where we accounted for the long-range hydrodynamic interaction between different beads, that is transmitted through the velocity field of the solvent [7]. For both models, the Langevin equations of motion in the mode space can be written in the form

$$\frac{d\vec{u}_\alpha}{dt} = -\Gamma_\alpha \vec{u}_\alpha + \vec{\zeta}_\alpha(t), \quad (2)$$

where $u_\alpha(t)$ is the amplitude of a normal mode α at time t, $\Gamma_\alpha = m\omega_\alpha^2\Lambda_\alpha$ is the mode relaxation rate, $\vec{\zeta}_\alpha(t)$ is thermal white noise that obeys the fluctuation-dissipation theorem

$$\langle\vec{\zeta}_\alpha(t)\vec{\zeta}_\beta(t')\rangle = 2k_BT\Lambda_\alpha\delta_{\alpha,\beta}\delta(t-t'), \quad (3)$$

and Λ_α is the mode mobility coefficient. The dependence of Λ_α on ω_α is sensitive to the hydrodynamic model in question. To account for both models in a single formula, we shall write the relaxation rate as $\Gamma_\alpha \simeq \bar{A}\omega_\alpha^\theta$, where (i) in the Rouse model $\theta = 2$, $\bar{A} = m/(3\pi\eta b)$ where η is the solvent viscosity and b is the bead diameter [3], and (ii) in the Zimm model $\theta = 2 - d_s + d_s/d_f$, $\bar{A} = Am/(6\pi\eta b\omega_o^{d_s/d_f-d_s})$, where A is a numerical constant [7].

We use Eq. (2) to evaluate the two-point correlation function for the time regime $\tau_0 \ll t \ll \tau_N$, where $\tau_0 = \bar{A}^{-1}\omega_0^{-\theta}$ and $\tau_N \simeq \bar{A}^{-1}\omega_0^{-\theta}N^{\theta/d_s}$ are the shortest and longest vibrational relaxation times (respectively). We obtain the following scaling form

$$\langle(\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0))^2\rangle = \frac{k_BT}{m\omega_o^{d_s}}(\bar{A}t)^\nu\Phi[|\vec{\ell}-\vec{\ell}'|/\ell(t)], \quad (4)$$

where $\Phi[v]$ is the scaling function [22,23] ($\Phi[0] = \text{const}$) and $\ell(t) = \omega_o^{d_s/d_l}(\bar{A}t)^{d_s/d_l\theta}$ is the (dimensionless) length

describing the propagation with distance, in topological space, of the bead-bead correlations or localized perturbations. In real space, this (dimensioned) propagation length is $\xi(t) \simeq b\ell(t)^{d_l/d_f} = b\omega_o^{d_s/d_f}\bar{A}^{\zeta}t^{\zeta}$, where $\zeta = d_s/(d_f\theta)$. Putting $\vec{\ell} = \vec{\ell}'$ in Eq. (4) it reduces simply to the (spatially averaged) MSD of a bead. Provided that $d_s < 2$, it shows the familiar *anomalous subdiffusion*

$$\langle \Delta\vec{u}(t)^2\rangle \equiv \langle(\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell},0))^2\rangle = Bt^{\nu}, \tag{5}$$

where $\nu = (2 - d_s)/\theta$ and B is a constant.

The scaling form Eq. (4) implies a crossover behavior around a time $t^*(r)$ for pairs at a distance $r \equiv |\vec{R}_{\ell} - \vec{R}_{\ell'}|$ apart, $t^*(r) = \bar{A}^{-1}\omega_0^{-\theta}(r/b)^{d_f\theta/d_s}$ [24]. For $t \ll t^*(r)$, the correlation function is very close to its static value

$$\langle(\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0))^2\rangle \approx \frac{k_BT}{m\omega_0^2}\left(\frac{r}{b}\right)^{d_w-d_f}, \tag{6}$$

where $d_w = 2d_f/d_s$. For longer times, $t^*(r) \ll t \ll \tau_N$, such that information has propagated much beyond the distance r, it approaches the MSD of a single bead, $\langle(\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0))^2\rangle \simeq Bt^{\nu}$, implying that the two beads are essentially moving together.

We note that the two-point correlation function Eq. (4) can be transformed from topological space to the real 3D Euclidean space,

$$\langle(\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0))^2\rangle = \frac{k_BT}{m\omega_0^2}\left(\frac{r}{b}\right)^{d_w-d_f}\Xi\left[\frac{t}{t^*(r)}\right]. \tag{7}$$

$\Xi[u]$ has the following asymptotes: (i) $\Xi[u] \simeq$ const for $u \ll 1$, and (ii) $\Xi[u] \sim u^{\nu}$ for $u \gg 1$. Equation (7) is particularly useful for the numerical analysis that we perform next.

To test the above analytic expression we first evaluate numerically the pair correlation function on a vibrating Sierpinski gasket obeying the Rouse dynamics [1,2,18]. In Fig. 1(a) we plot, on a log-log scale, four averaged two-point correlation functions versus the normalized time t/τ_0 [25]. Note the crossover from a constant value, that increases with increasing r as predicted by Eq. (6), to an anomalous subdiffusion time regime, identical to that of the single particle MSD (effectively equal to the correlation function for $r = b = 5$ Å). In the subdiffusion regime, the behavior is essentially independent of the distance r. Also note that the crossover time increases with increasing r as implied by the scaling with r of $t^*(r)$. A normalized version of Fig. 1(a), shown in Fig. 1(b), asserts the validity of Eq. (7).

We now turn to calculate the fractal DSF with rotational and translational motion arrested,

$$S(\vec{k},t) \equiv \frac{1}{N}\left\langle\sum_{\vec{\ell},\vec{\ell}'} e^{i\vec{k}\cdot(\vec{R}(\vec{\ell},t)-\vec{R}(\vec{\ell}',0))}\right\rangle, \tag{8}$$

where $\vec{R}(\vec{\ell})$ is the coordinate of a node $\vec{\ell}$ in the center of mass coordinate frame and the sums run over all network nodes. Changing to displacement variables, $\vec{R}(\vec{\ell},t) = \vec{R}_{\rm eq}(\vec{\ell}) + \vec{u}(\vec{\ell},t)$, using the Gaussian property of the stochastic variable $\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)$ [11] and the isotropy of the scalar elasticity model, and performing angular averaging, we find, omitting from now on the subscript "eq" in $\vec{R}_{\rm eq}(\vec{\ell})$,

$$S(k,t) = \frac{1}{N}\sum_{\vec{\ell},\vec{\ell}'}\frac{\sin[kR_{\ell\ell'}]}{kR_{\ell\ell'}}e^{-(k^2/6)\langle(\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0))^2\rangle}, \tag{9}$$

where $R_{\ell\ell'} = |\vec{R}(\vec{\ell}) - \vec{R}(\vec{\ell}')|$ is the Euclidean, real-space, distance between beads $\vec{\ell}$ and $\vec{\ell}'$.

At short times, $t \ll t^*(k^{-1}) \sim k^{-d_f\theta/d_s}$, information did not have time to negotiate a "blob" of linear size $\sim k^{-1}$ and we find that the DSF did not decay much, $S(k,t) \simeq S(k)$. At longer times, $t^*(k^{-1}) \ll t \ll \tau_N$, i.e., when $1 \ll k\xi(t) \ll kR_g$, information has propagated beyond the scattering wavelength $\sim 1/k$. Physically, this implies that the blob of size $\sim 1/k$, that is controlling the relaxation at wave vector $\vec{k}$, is now moving almost coherently as if it was a single bead. At this time regime we find a *stretched exponential* decay of the DSF,

$$S(k,t) \approx S(k)\exp[-(\Gamma_k t)^{\nu}], \tag{10}$$

where

$$\Gamma_k = (B/6)^{1/\nu}k^{2/\nu}. \tag{11}$$

Note that the stretching exponent is exactly the anomalous diffusion exponent ν. The stretched exponential decay, together with the dependence of the stretching exponent ν on the broken dimensions d_s and d_f, is thus a strong signature of the fractal structure. Corrections due to rotational and translational diffusion are considered elsewhere and are shown to vanish for large fractals [23]. In Fig. 2 we plot the DSF (and SF) for the Sierpinski gasket.

The anomalous wave number dependence of the relaxation rate, $\Gamma_k \sim k^{2/\nu}$, and the anomalous diffusion, $\langle\Delta\vec{u}(t)^2\rangle \sim t^{\nu}$, can be explained using simple scaling hypotheses. For the relaxation rate we assume $\Gamma_k = Dk^2h(k\bar{u})$, where $h(x)$ is a scaling function and D is the center of mass diffusion coefficient of the fractal. We make use of the generalized Landau-Peierls instability, $\bar{u} \sim N^{1/d_s-1/2}$, and take $D \sim N^{-1}$ and $D \sim R_g^{-1} \sim N^{-1/d_f}$ for the Rouse- and Zimm-type models of friction [11], respectively. Demanding that Γ_k is independent of N for $k\bar{u} \gg 1$, the scaling function for $x \gg 1$ must satisfy $h(x) \sim x^{2d_s/(2-d_s)}$ (for the Rouse model) and $h(x) \sim x^{2d_s/[d_f(2-d_s)]}$ (for the Zimm model), leading to $\Gamma_k \sim k^{2/\nu}$ with $\nu = 1 - d_s/2$ (Rouse) and $\nu = (2 - d_s)/(2 - d_s + d_s/d_f)$ (Zimm) as stated. Similarly, for the MSD we assume $\langle\Delta\vec{u}(t)^2\rangle = \bar{u}^2y(t/\tau_N)$, where $y(x)$ is the scaling function and τ_N is the longest vibrational relaxation time. Assuming $\tau_N \simeq \bar{u}^2/D$, such that $\tau_N \sim N^{2/d_s}$ (Rouse) and

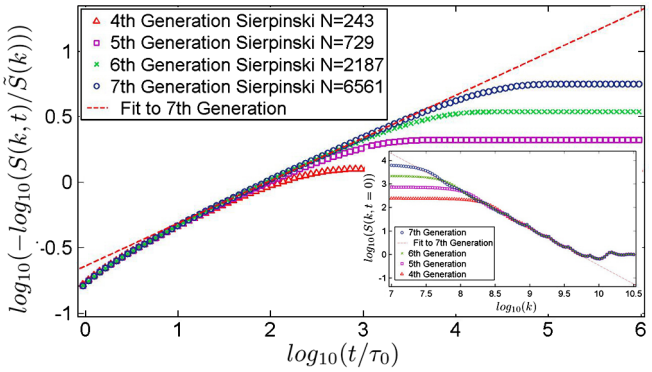


FIG. 2 (color online). $\log_{10}\{-\log_{10}[S(k,t)/\tilde{S}(k)]\}$ for $k = 10^{10}$ m^{-1} is plotted versus $\log_{10}(t/\tau_0)$ for vibrating Sierpinski gaskets of various sizes (Rouse model) [25]. Here $\tilde{S}(k)$ is the frozen SF [Eq. (9) with $e^{(\cdots)}$ set to unity]. Note that a straight line, whose slope is positive, is formed on an intermediate time window that widens up as we move from smaller to larger gaskets, demonstrating the diminishing contribution of finite size effects. The stretching exponent obtained from the plot, 0.325, is in excellent agreement with the theoretical value $\nu \simeq 0.317$. Inset: The static SF, $S(k) \equiv S(k, t = 0)$, against k on a log-log scale. A clear power law decay, with an exponent of -1.583, is visible for all gaskets, in excellent agreement with theory [2,12] $\tilde{S}(k) \sim k^{-d_f}$, $d_f \simeq 1.585$. In contrast to $\tilde{S}(k)$, $S(k)$ does account for the contribution of thermal vibrations. However, as is evident from the plot, vibrations have a negligible effect on the static SF for the chosen parameter values [25] and $S(k) \simeq \tilde{S}(k)$.

$\tau_N \sim N^{2/d_s-1+1/d_f}$ (Zimm), and demanding that for $t \ll \tau_N$ the MSD is independent of N, it follows that $y(x) \sim x^\nu$ for $x \ll 1$ (with ν taking the above stated values associated with the Rouse and Zimm models) leading to $\langle \Delta\vec{u}(t)^2 \rangle \sim t^\nu$.

In this Letter we presented a thorough study of the DSF $S(k, t)$ of vibrating fractals in solutions. Our main result is that at large wave numbers the decay of the DSF is strongly influenced by the anomalous diffusion of the spatially averaged MSD of a network bead. The result for large fractals is a stretched exponential decay of the DSF, which reduces to the known DSF of a linear Rouse polymer chain ($d_s = 1$), and a linear-Gaussian Zimm polymer chain ($d_s = 1$, $d_f = 2$) [11]. Among possible applications are the following. (i) Proteins, for which a stretched exponential decay has been recently measured by neutron spin echo [16], supporting their fractal-like structure. (ii) Glass forming colloidal suspensions [17]. Assuming analogy to 3D percolation network, that implies $d_f \simeq 2.48$ and $d_s \simeq 1.328$, and accounting for the hydrodynamic coupling, we suggest a stretched exponential decay with $\nu \simeq 0.556$, remarkably close to the observed value $\nu \simeq 0.6$. (iii) Colloidal gels [26], that show a clear fractal structure and for which a Zimm-type dynamics and bond-bending potential explain well the observed stretched exponential decay [7]. (iv) Chromatin [14,15], for which it was recently shown that telomeres perform anomalous subdiffusion with $\nu \simeq 0.32$ [27]. This may be interpreted, within the Rouse model that yields $\nu = 1 - d_s/2$ (noting that in such a dense polymer system hydrodynamics is likely to be screened), by $d_s \simeq 1.36$. This value of d_s is remarkably close to that of percolation clusters in $2 < d < 5$ dimensions and suggests the presence of DNA cross-links (e.g., via ligation). According to the present calculation, it is suggested that the DSF of chromatin will decay as a stretched exponential with stretching exponent $\simeq 0.32$, which can motivate experiments in this direction. We believe that our results can also be applied to a variety of other systems exhibiting fractal structure.

R. G. and J. K. acknowledge support from the Deutsch-Israelische Projektkooperation (DIP). R. G. acknowledges partial support from the National Science Foundation under Grant No. PHY05-51164. S. R. acknowledges support from the Converging Technologies program of the Israeli Council for higher education.

*Corresponding author.
rgranek@bgu.ac.il

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature* (W. H. Freeman, New York, 1982).
[2] For a review, see T. Nakayama, K. Yakubo, and R. L. Orbach, Rev. Mod. Phys. **66**, 381 (1994).
[3] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[4] M. B. Enright and D. M. Leitner, Phys. Rev. E **71**, 011912 (2005).
[5] R. Burioni *et al.*, Proteins **55**, 529 (2004); R. Burioni *et al.*, Europhys. Lett. **58**, 806 (2002).
[6] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008); M. de Leeuw *et al.*, PLoS ONE **4**, e7296 (2009); S. Reuveni, R. Granek, and J. Klafter, Proc. Natl. Acad. Sci. U.S.A. **107**, 13 696 (2010).
[7] R. Granek, Phys. Rev. E **83**, 020902(R) (2011).
[8] For a review, see A. Banerji and I. Ghosh, Cell Mol. Life Sci. **68**, 2711 (2011).
[9] M. E. Cates, J. Phys. (Paris) **46**, 1059 (1985); J. E. Martin, D. Adolf, and J. P. Wilcoxon, Phys. Rev. A **39**, 1325 (1989); A. G. Zilman and R. Granek, Phys. Rev. E **58**, R2725 (1998).
[10] For a review, see C. M. Sorensen, Aerosol Sci. Technol. **35**, 648 (2001).
[11] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
[12] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, England, 1995).
[13] E. L. Aiden *et al.*, Science **326**, 289 (2009).
[14] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. (Paris) **49**, 2095 (1988); A. Y. Grosberg *et al.*, Europhys. Lett. **23**, 373 (1993).
[15] J. G. McNally and D. Mazza, EMBO J. **29**, 2 (2010); A. J. Einstein *et al.*, J. Pathol. Bacteriol. **185**, 366 (1998); V. Bedin *et al.*, BMC Cancer **10**, 260 (2010).

[16] J. Lal *et al.*, J. Mol. Biol. **397**, 423 (2010).
[17] J. Mattsson *et al.*, Nature (London) **462**, 83 (2009).
[18] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992), 2nd ed.
[19] A. Blumen *et al.*, Macromolecules **37**, 638 (2004); A. Blumen *et al.*, Phys. Rev. E **67**, 061103 (2003).
[20] Although we constantly use the polymer physics terms "Rouse" and "Zimm" models, these terms are only used here to emphasize the hydrodynamic friction model in question, while the actual models are quite different from their polymer analogs.
[21] J. W. Kantelhardt and A. Bunde, Phys. Rev. E **56**, 6693 (1997).
[22] $\Phi(v) = 6d_s \int_0^\infty dz z^{d_s-3}(1 - f[z^{d_s/d_l}v]e^{-z^\theta})$, where $f(x)$ is the disorder averaged mode scaling function [3,21]; $\Phi[0] = 6d_s\Gamma[d_s/2]/(2 - d_s)$ (Rouse model) and $\Phi[0] = 6d_s\Gamma[1/(d_w - d_f + 1)]/(2 - d_s)$ (Zimm model).
[23] S. Reuveni, J. Klafter, and R. Granek, Phys. Rev. E **85**, 011906 (2012).
[24] Note that $t^*(R_g) \simeq \tau_N$.
[25] We use $b = 5 \times 10^{-10}$ m, $m = 4 \times 10^{-25}$ kg, $\omega_0 = 10^{12}$ s^{-1}, $\eta = 8.94 \times 10^{-4}$ Pa s, $T = 298$ K, yielding $\tau_0 \simeq 10.53$ ps. The size $N = 6561$ corresponds to $R_g \simeq 42.6$ nm and $\tau_N \simeq 4.11$ μs. These parameter values are similar to those in proteins, except for N which is purposely exaggerated [23].
[26] A. H. Krall and D. A. Weitz, Phys. Rev. Lett. **80**, 778 (1998).
[27] I. Bronstein *et al.*, Phys. Rev. Lett. **103**, 018102 (2009).

Dynamic structure factor of vibrating fractals: Proteins as a case study

Shlomi Reuveni,[1,2] Joseph Klafter,[1] and Rony Granek[3,*]
[1]*School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel*
[2]*Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel*
[3]*Stella and Avram Goren-Goldstein Department of Biotechnology Engineering, Ilse Katz Institute for Meso and Nanoscale Science and Technology, and Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering, Ben-Gurion University of The Negev, Beer Sheva 84105, Israel*
(Received 18 September 2011; published 10 January 2012)

We study the dynamic structure factor $S(k,t)$ of proteins at large wave numbers k, $kR_g \gg 1$, where R_g is the gyration radius. At this regime measurements are sensitive to internal dynamics, and we focus on vibrational dynamics of folded proteins. Exploiting the analogy between proteins and fractals, we perform a general analytic calculation of the displacement two-point correlation functions, $\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle$. We confront the derived expressions with numerical evaluations that are based on protein data bank (PDB) structures and the Gaussian network model (GNM) for a few proteins and for the Sierpinski gasket as a controlled check. We use these calculations to evaluate $S(k,t)$ with arrested rotational and translational degrees of freedom, and show that the decay of $S(k,t)$ is dominated by the spatially averaged mean-square displacement of an amino acid. The latter has been previously shown to evolve subdiffusively in time, $\langle[\vec{u}_i(t)-\vec{u}_i(0)]^2\rangle \sim t^{\nu}$, where ν is the anomalous diffusion exponent that depends on the spectral dimension d_s and fractal dimension d_f. As a result, for wave numbers obeying $k^2\langle\vec{u}^2\rangle \gtrsim 1$, $S(k,t)$ effectively decays as a stretched exponential $S(k,t) \simeq S(k)e^{-(\Gamma_k t)^{\beta}}$ with $\beta \simeq \nu$, where the relaxation rate is $\Gamma_k \sim (k_B T/m\omega_o^2)^{1/\beta}k^{2/\beta}$, T is the temperature, and $m\omega_o^2$ the GNM effective spring constant describing the interaction between neighboring amino acids. The static structure factor is dominated by the fractal character of the native fold, $S(k) \sim k^{-d_f}$, with negligible to marginal influence of vibrations. The analytical expressions are first confronted with numerically based calculations on the Sierpinski gasket, and very good agreement is found between simulations and theory. We then perform PDB-GNM-based numerical calculations for a few proteins, and an effective stretched exponential decay of the dynamic structure factor is found, albeit their relatively small size. However, when rotational and translational diffusion are added, we find that their contribution is never negligible due to finite size effects. While we can still attribute an effective stretching exponent β to the relaxation profile, this exponent is significantly larger than the anomalous diffusion exponent ν. We compare our theory with recent neutron spin-echo studies of myoglobin and hemoglobin and conclude that experiments in which the rotational and translational degrees of freedom are arrested, e.g., by anchoring the proteins to a surface, will improve the detection of internal vibrational dynamics.

DOI: 10.1103/PhysRevE.85.011906 PACS number(s): 87.15.ad, 05.40.−a, 47.53.+n

I. INTRODUCTION

In the last few years there has been renewed interest in the fractal-like nature of natively folded proteins [1–13]. In analogy with mathematically constructed fractals, it has been shown that each protein can be associated with characteristic broken dimensions. Diversity was understood in terms of the unique exponents that characterize each protein [3,5,8]. This viewpoint has allowed description of protein dynamics on a universal level and has led to a unified approach toward observed anomalies in the vibrational dynamics of proteins. On the experimental side, evidence of fractality in proteins came from electron-spin relaxation measurements [11] and neutron scattering [12]. Indirect evidence came from single-molecule experiments that have reported anomalous behaviors involving power laws in time [14] and from molecular dynamics simulations that have shown anomalous diffusion of vibrational energy [3], dihedral angels [15], and amino acids [16]. These observations were further backed by various elastic network-based studies [5–8,10] concluding that anomalies in the vibrational dynamics of proteins are a consequence of a fractal-like structure [8]. Biological relevance of the fractal properties of proteins was discussed in Refs. [6–8,13]. More recently, the fractal analysis has been proven useful in explaining the action of antibodies [17].

Fractals are characterized by a few broken dimensions [18,19]: (1) the mass fractal dimension d_f that governs the scaling $M(r) \sim r^{d_f}$ of the mass $M(r)$ enclosed in concentric spheres of radius r, (2) the spectral dimension d_s that governs the scaling $g(\omega) \sim \omega^{d_s-1}$ of the vibrational density of states (DOS) $g(\omega)$ with frequency ω [19–21], and (3) the topological dimension d_l that governs the scaling $M(l) \sim l^{d_l}$ of the mass $M(l)$ enclosed in concentric "spheres" of radius l in the topological (or "manifold" ["chemical"]) space. One may also define, instead of d_l, the chemical length (or minimal path) dimension $d_{\min} = d_f/d_l$ that relates the real space distance r between two points on the fractal to the minimal path distance l between these points along the fractal network links, $l \sim r^{d_{\min}}$. The dimensions d_f and d_s have been computed for a large number of proteins using the native fold structures obtained from the protein data bank [7]. The fractal dimension d_f is computed from these structures in a straightforward manner. The topological dimension d_l has also been computed and found close to the fractal dimension d_f [22]. The computation of d_s requires a network elasticity model, and the Gaussian

*rgranek@bgu.ac.il

network model (GNM) has been mostly used [23]. The spectral dimension of the vast majority of proteins has been found to be smaller than two [4,6,7]. Importantly, it leads to the generalized Landau-Peierls instability, $\langle u^2 \rangle \sim N^{\frac{2}{d_s}-1}$ where u is an amino acid displacement (averaged over all amino acids) and N is the number of amino acids [4,6,24]. By invoking marginal stability, which allows proteins to attain maximum fluctuations (or "flexibility") but keep their native fold structure, a universal equation of state that relates d_s, d_f, and N, for all natively folded proteins [6], has been deduced. The equation has been validated for about 5000 proteins, and remarkable agreement has been found, regardless of protein source or function [7].

Recent advances in high-resolution inelastic neutron scattering, available with neutron spin-echo (NSE) spectroscopy, have turned this approach useful in the analysis of biomolecule flexibility and vibrational dynamics [25,26]. NSE studies that measure the dynamic structure factor $S(k,t)$ have been recently performed on horse heart myoglobin and bovine hemoglobin in solutions [27]. In the large wave number k regime corresponding to $kR_g \gg 1$, where R_g is the gyration radius, and at low concentrations and times shorter than 1 ns, the result is a stretched exponential relaxation, $S(k,t) \sim e^{-(\Gamma_k t)^\beta}$ with $\beta \simeq 0.4 \pm 0.03$ for both proteins and independent of k. The relaxation rate has been found to scale as $\Gamma_k \sim k^{2/\beta}$. Inelastic neutron scattering, a complementary method to NSE, is used for exploring protein dynamics at high frequencies ("short times") [28,29]. Inelastic neutron-scattering experiments and MD simulations performed on lysozyme showed a non-Lorentzian spectra that corresponds to a nonexponential decay with time of the dynamic structure factor [30]. These findings are in accord with the NSE findings described above and with the theory developed herein.

Phenomenologically, the nonexponential relaxation of $S(k,t)$ can be approached by assuming that collective (averaged) observables, associated with the internal dynamics of proteins, follow fractional Brownian motion rather than regular Brownian motion [31,32]. The assumption is justified *a posteriori* by comparison to MD simulations and experiments. Another approach toward the anomalous relaxation of $S(k,t)$ is based on an analogy with polymer dynamics theory, and one conjectures that $S(k,t)$ decays mainly due to the time evolution of the amino acid mean-square displacement (MSD) [33–35]. Combined with the above experimental observations, this conjecture suggests that the MSD evolves anomalously in time, i.e., as $\sim t^\nu$ with $\nu \simeq \beta$. Yet the anomalous diffusion exponent does not fit any of the polymer theory exponents, $2/3$ in the Zimm model and $1/2$ in the Rouse model. More importantly, a folded protein is clearly nothing like a solvated, open, flexible polymer, which fluctuates between its many available configurations without having any underlying scaffold. Hence, the use of polymer theory can only serve as a guiding tool.

In this paper we take a first-principles approach and relate the anomalies in the decay of the dynamic structure factor with the fractal-like nature of proteins. We first note that on time scales less than or of order nanosecond a protein is not expected to experience any unfolding-refolding dynamics, even on a local scale, and all dynamics is expected to be associated with (overdamped) vibrations about the folded structure. Due to the protein fractal-like structure, the MSD of an amino acid, averaged over all amino acids of the protein, has been shown to be subdiffusive, that is, to scale at short times as $\sim t^\nu$. The exponent ν depends on the fractal and spectral dimensions, $\nu = 1 - d_s/2$ in a Rouse-type model [5,8,35], $\nu = (2 - d_s)/(2 - d_s + d_s/d_f)$ in a Zimm-type model [10], and $\nu = 2 - d_s$ for vanishing friction [5]. Note that although we constantly use the polymer physics terms "Rouse" and "Zimm," these terms are used only in order to emphasize the hydrodynamic friction model in question. In a Rouse-type model the friction is local, while in a Zimm-type model the friction is long range due to hydrodynamic interactions. In contrast to the original, polymeric, use of the terms "Rouse" and "Zimm," we obviously do not assume that a protein is modeled by a one-dimensional Gaussian chain. On the contrary, as described later, the native three-dimensional structure is most definitely taken into consideration. Hence the anomalous diffusion exponents are in general different from their polymer values $\nu = 1/2$ (Rouse) and $\nu = 2/3$ (Zimm) [33]. Special fractals may yield the linear polymer Zimm exponent $\nu = 2/3$, e.g., the Vicsek fractals studied numerically by Blumen and co-workers [35] (as can also be verified by using the above analytic expression for ν); however, this is not typical. For further discussion about the difference between the Zimm-like and Rouse-like models used here, and those used in polymer theory, the reader is referred to Ref. [10].

Since all amino acids contribute to the dynamic structure factor, it should indeed exhibit the spatially averaged MSD. However, this single amino acid picture fails to capture both static and dynamic correlations between amino acids; e.g., it is unable to predict the static structure factor $S(k)$, and so this simplified viewpoint may not be complete. We thus make a complete calculation of the dynamic structure factor based on the fractal analogy and delineate the regimes for which the decay of $S(k,t)$ is dominated by the MSD, thus connecting the fractal exponents d_s and d_f to the stretched exponential decay of $S(k,t)$. We also predict the dependence of the relaxation rate on the effective spring constant describing the interaction between neighboring amino acids. Moreover, we add the nontrivial effect of rotational diffusion on the relaxation profile. In addition, we demonstrate the signature of protein fractal properties and fractal controlled vibrations on the static structure factor.

In order to obtain the dynamic structure factor, we calculate first the vibrational amino acid pair correlation function $\langle [u_i(t) - u_j(0)]^2 \rangle$, which is a fundamental ingredient of the dynamic structure factor. We expand this correlation function in normal modes, obtain its scaling form, and relate it to two dynamical quantities that we have focused on in the past: (1) the MSD of an amino acid, and (2) the autocorrelation function of the distance between a pair of amino acids. We dwell on the different asymptotes of the pair correlation function that are relevant for the decay profile of the dynamic structure factor. We show that above a crossover time that signifies the passage of information (or energy [3]) between the two residues, the pair correlation function in question approaches the MSD. As the latter is subdiffusive, it leads to the stretched exponential decay of the dynamic structure factor.

II. MODEL DEFINITIONS

We repeat briefly the model definitions and assumptions, following the notations of Ref. [5]. Protein vibrations are discussed using the Gaussian network model (GNM) [20,23]. The model assigns identical springs between α-carbon pairs that are distant less than a cutoff distance R_c, whose typical values range between 6 to 8 Å. Each α-carbon, henceforth named a "bead," is assigned an average amino acid mass. In what follows, we assume that the network forms a disordered fractal. The index of a bead, or its coordinate in topological space, is denoted symbolically by the "vector" $\vec{l}$. The vector $\vec{R}(\vec{l})$ denotes its position in real space. (Note that this indexing method is equivalent to the indexing method used in the Abstract and Sec. I. In what follows we will use the two indexing methods interchangeably.) The ground configurational state of the protein is described by the set of coordinates $\vec{R}_{\rm eq}(\vec{l})$, and deviations from the ground state are denoted by the displacements $\vec{u}(\vec{l}) = \vec{R}(\vec{l}) - \vec{R}_{\rm eq}(\vec{l})$. The GNM Hamiltonian is

$$H[\{\vec{u}(\vec{l})\}] = \frac{1}{2} m\omega_o^2 \sum_{\langle \vec{l}\vec{l}'\rangle} [\vec{u}(\vec{l}) - \vec{u}(\vec{l}')]^2, \tag{1}$$

where $\langle \vec{l}\,\vec{l}'\rangle$ stands for pairs connected by springs, ω_o is the spring self-frequency, and m is the bead mass ($m\omega_o^2$ is the spring constant). The eigenstates (normal modes) $\Psi_\alpha(\vec{l})$ of the Hamitonian (1), are solutions of the eigenvalue equation

$$\omega_o^2 \sum_{\vec{l}'\in\vec{l}} [\Psi_\alpha(\vec{l}') - \Psi_\alpha(\vec{l})] = -\omega_\alpha^2 \Psi_\alpha(\vec{l}), \tag{2}$$

where ω_α is the mode frequency. Here $\vec{l}' \in \vec{l}$ denotes beads connected by springs to the bead $\vec{l}$. $\{\Psi_\alpha(\vec{l})\}$ form an orthonormal set [19,21] such that

$$\sum_{\vec{l}} \Psi_\alpha(\vec{l})\Psi_\beta(\vec{l}) = \delta_{\alpha,\beta} \tag{3}$$

and

$$\sum_{\alpha} \Psi_\alpha(\vec{l})\Psi_\alpha(\vec{l}') = \delta_{\vec{l},\vec{l}'}. \tag{4}$$

This allows to define a normal mode transform

$$\vec{u}_\alpha = \sum_{\vec{l}} \vec{u}(\vec{l})\Psi_\alpha(\vec{l}), \tag{5}$$

and an inverse transform

$$\vec{u}(\vec{l}) = \sum_{\alpha} \vec{u}_\alpha \Psi_\alpha(\vec{l}), \tag{6}$$

where $\vec{u}_\alpha$ is the amplitude of the normal mode $\Psi_\alpha(\vec{l})$. In the normal mode "space," the Hamiltonian is diagonal,

$$H[\{\vec{u}_\alpha\}] = \frac{1}{2} m \sum_{\alpha} \omega_\alpha^2 \vec{u}_\alpha^2. \tag{7}$$

The equipartition theorem then dictates that at thermal equilibrium

$$\langle \vec{u}_\alpha \cdot \vec{u}_\beta \rangle_T = \frac{3k_B T}{m\omega_\alpha^2}\delta_{\alpha,\beta}. \tag{8}$$

On a fractal, the normal modes $\Psi_\alpha(\vec{l})$ are strongly localized in space, unlike the oscillatory behavior characteristic of uniform networks. A disorder averaged eigenstate can be defined according to

$$\bar{\Psi}(\omega_\alpha, |\vec{\ell} - \vec{\ell}'|) = N\langle \Psi_\alpha(\vec{\ell})\Psi_\alpha(\vec{\ell}')\rangle_{\rm dis}, \tag{9}$$

where $\langle\cdots\rangle_{\rm dis}$ denotes disorder averaging, i.e., averaging over all realizations of the fractal keeping the nodes $\vec{\ell}$ and $\vec{\ell}'$ fixed, or averaging, within a given realization, over all different pair of nodes $\vec{\ell}$ and $\vec{\ell}'$ that have the same topological space distance $|\vec{\ell} - \vec{\ell}'|$. Note that mode normalization implies $\langle \Psi_\alpha(\vec{\ell})^2\rangle_{\rm dis} = 1/N$. It has been shown that $\bar{\Psi}(\omega_\alpha, l)$ obeys the following scaling form [18,20,21]:

$$\bar{\Psi}(\omega_\alpha, l) = f[(\omega_\alpha/\omega_o)^{d_s/d_l} l], \tag{10}$$

where $f(y)$ is the scaling function. For $y \gg 1$, $f(y)$ is *exponentially decaying*, and, for $y \ll 1$, $f(y) \simeq 1 - C_0 \times y^2$, where C_0 is a numerical constant [21,36,37].

III. TWO-POINT CORRELATION FUNCTION AND MEAN-SQUARE DISPLACEMENT

A. Introduction

In order to derive the dynamic structure factor of proteins at large wave numbers and short times, we discuss first the relevant displacement pair correlation functions that are needed for this calculation. We are specifically interested in the pair correlation function $\langle[\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2\rangle$, where it is understood that spatial averaging, i.e., averaging over many origins $\vec{\ell}'$ that are sufficiently far from the periphery, has been performed, thus making this correlation function depend only on the relative separation $|\vec{\ell} - \vec{\ell}'|$ in topological space. In particular, for $\vec{\ell} = \vec{\ell}'$, this correlation function reduces simply to the (spatially averaged) MSD of an amino acid.

B. Basic dynamics

In Refs. [5,8,10] we derived the normal mode space Langevin equations for the fractal-like protein in the high damping and vanishing damping limits. In the high damping limit, which is our main focus here due to its relevance to protein dynamics in solutions, two models were considered: (1) a Rouse-type model in which the hydrodynamic friction is local, and (2) a Zimm-type model where we accounted for the hydrodynamic interaction between different amino acids, that is, transmitted through the velocity field of the solvent. For both models, the Langevin equations of motion in the mode space can be written in the form

$$\frac{d\vec{u}_\alpha}{dt} = -\Gamma_\alpha \vec{u}_\alpha + \vec{\zeta}_\alpha(t), \tag{11}$$

where $u_\alpha(t)$ is the amplitude of a normal mode α at time t,

$$\Gamma_\alpha = m\omega_\alpha^2 \Lambda_\alpha \tag{12}$$

is the mode relaxation rate, $\vec{\zeta}_\alpha(t)$ is thermal white noise that obeys the fluctuation-dissipation theorem

$$\langle \vec{\zeta}_\alpha(t)\vec{\zeta}_\beta(t')\rangle = 2k_B T \Lambda_\alpha \delta_{\alpha,\beta}\delta(t - t'), \tag{13}$$

and Λ_α is the mode mobility coefficient. Using Eq. (11), the time autocorrelation function of a mode amplitude obeys a simple exponential decay controlled by Γ_α,

$$\langle \vec{u}_\alpha(t)\cdot\vec{u}_\alpha(0)\rangle = \langle \vec{u}_\alpha^{\,2}\rangle_T \mathrm{e}^{-\Gamma_\alpha t}. \tag{14}$$

The dependence of Λ_α on ω_α is sensitive to the hydrodynamic model in question. In the Rouse model

$$\Lambda_\alpha = 1/m\gamma, \tag{15}$$

independent of frequency, where $m\gamma$ is the friction coefficient of a bead. Thus $m\gamma \simeq 3\pi\eta b$ for Stokes friction, where b is the bead diameter, taken for simplicity equal to the mean bond length ($b \leqslant R_c$). In the Zimm model,

$$\Lambda_\alpha = (A/6\pi\eta b)(\omega_\alpha/\omega_o)^{\frac{d_s}{d_f}-d_s}, \tag{16}$$

where A is a numerical constant,

$$A = \frac{d_l \pi^{d_l/2}}{\Gamma[d_l/2+1]}\int_0^\infty dx\, x^{d_l-1-d_l/d_f} f(x). \tag{17}$$

[$f(x)$ is the mode scaling function defined in Eq. (10).]

To account for both models in a single formula, we shall write the relaxation rate as

$$\Gamma_\alpha \simeq \bar{A}\,\omega_\alpha^{\theta} \tag{18}$$

where (1) in the Rouse model

$$\theta = 2\,; \quad \bar{A} = 1/\gamma = m/(3\pi\eta b), \tag{19}$$

and (2) in the Zimm model

$$\theta = 2 - d_s + d_s/d_f; \quad \bar{A} = A\, m/\left(6\pi\eta b \omega_o^{\frac{d_s}{d_f}-d_s}\right). \tag{20}$$

C. Two-point correlation function: Basic expressions

Expanding the two-point correlation function in terms of the exact normal modes $\Psi_\alpha(\vec{\ell})$, we find

$$\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle = \sum_\alpha \langle \vec{u}_\alpha^{\,2}\rangle_T [\Psi_\alpha(\vec{\ell})^2+\Psi_\alpha(\vec{\ell}')^2-2\Psi_\alpha(\vec{\ell})\Psi_\alpha(\vec{\ell}')\mathrm{e}^{-\Gamma_\alpha t}]. \tag{21}$$

Performing disorder average over Eq. (21) [see definition after Eq. (9)], and using the definition of the disorder averaged normal mode $\bar{\Psi}(\omega_\alpha,\ell)$, Eq. (9), the two-point correlation function that we focus on is found to be

$$\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle = \frac{2}{N}\sum_\alpha \langle \vec{u}_\alpha^{\,2}\rangle_T [1-\bar{\Psi}(\omega_\alpha,|\vec{\ell}-\vec{\ell}'|)\mathrm{e}^{-\Gamma_\alpha t}]. \tag{22}$$

If the fractal (protein) is sufficiently large, the frequency spectrum is dense, and we can approximate the sum by an integral over the frequency ω using the DOS $g(\omega) = n_o\omega^{d_s-1}$, where $n_o = N d_s/\omega_o^{d_s}$ is chosen such that $\int_0^{\omega_o} d\omega g(\omega) = N$. Making use of the equipartition theorem, Eq. (8), we obtain

$$\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle = 6d_s\frac{k_BT}{m\omega_o^{d_s}}\int_{\omega_{\min}}^{\omega_o} d\omega\omega^{d_s-3}\left[1-\bar{\Psi}(\omega,|\vec{\ell}-\vec{\ell}'|)\mathrm{e}^{-\bar{A}\omega^\theta t}\right]. \tag{23}$$

Note that the lower and upper integration limits in Eq. (23) set the shortest and longest vibrational relaxation times in the system. The longest relaxation time τ_N is the inverse of the smallest relaxation rate leading to $\tau_N = \Gamma(\omega_{\min})^{-1} \simeq \bar{A}^{-1}\,\omega_{\min}^{-\theta}$, where $\omega_{\min} \simeq \omega_o(R_g/b)^{-d_f/d_s}$, i.e.,

$$\tau_N \simeq \bar{A}^{-1}\,\omega_o^{-\theta}(R_g/b)^{\theta d_f/d_s} \sim N^{\theta/d_s}. \tag{24}$$

The shortest relaxation time τ_0 is the inverse of the largest relaxation rate, thereby

$$\tau_0 = \Gamma(\omega_o)^{-1} = \bar{A}^{-1}\,\omega_o^{-\theta}. \tag{25}$$

Focusing on the intermediate time regime $\tau_0 \ll t \ll \tau_N$ we can set the lower and upper limits of integration in Eq. (23) to 0 and infinity, respectively. Changing the variable of integration from ω to z, where $z^\theta = \bar{A}\omega^\theta t$, and using the scaling form of the eigenstates, $\Psi(\omega,l) = f[(\omega_\alpha/\omega_o)^{d_s/d_l}l]$, we obtain the following scaling form:

$$\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle = \frac{k_BT}{m\omega_o^{d_s}}(\bar{A}t)^\nu\ \Phi[|\vec{\ell}-\vec{\ell}'|/\ell(t)], \tag{26}$$

where

$$\Phi(v) = 6d_s\int_0^\infty dz z^{d_s-3}\left(1-f[z^{d_s/d_l}v]\mathrm{e}^{-z^\theta}\right) \tag{27}$$

and where $\ell(t) = \omega_o^{d_s/d_l}(\bar{A}t)^{\frac{d_s}{d_l\theta}}$ is the (dimensionless) length describing the propagation with distance, in topological space, of the bead-bead correlations or force (energy) perturbations. In real space, this (dimensioned) propagation length is

$$\xi(t) \simeq b\ell(t)^{d_l/d_f} = b\,\omega_o^{d_s/d_f}\,\bar{A}^\zeta\, t^\zeta, \tag{28}$$

where

$$\zeta = \frac{d_s}{d_f\theta} \tag{29}$$

is the real space propagation length exponent. We note that in the Rouse model $\zeta = 1/d_w$, where d_w, the random walk (RW) anomalous diffusion exponent, is given by the Alexander-Orbach relation [20]

$$d_w = \frac{2d_f}{d_s} \tag{30}$$

[i.e. $\langle r^2(t)\rangle \sim t^{2/d_w}$ for a random walk on the same network].

Molecular dynamics simulations aimed at characterizing the spread of vibrational energy in proteins [3] and random-walk simulations on proteins [8] have confirmed the power-law behavior of the propagation length both directly and indirectly. In particular, it was found that in proteins $d_w > 2$; i.e., propagation of correlations and force (energy) perturbations in the Rouse model is subdiffusive in time.

D. Mean-square displacement

In particular, for $\vec{\ell} = \vec{\ell}'$ we recover the previously derived vibrational MSD of a bead, averaged over all network beads. Provided that $d_s < 2$, it shows the familiar *anomalous subdiffusion*

$$\langle \Delta\vec{u}(t)^2\rangle \equiv \langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell},0)]^2\rangle = B\,t^\nu. \tag{31}$$

The exponent ν is

$$\nu = (2 - d_s)/\theta, \tag{32}$$

yielding

$$\nu = 1 - \frac{d_s}{2} \tag{33}$$

in the Rouse model, and

$$\nu = \frac{2 - d_s}{2 - d_s + d_s/d_f} = \frac{d_w - d_f}{d_w - d_f + 1} \tag{34}$$

in the Zimm model. The prefactor B is

$$B = \Phi[0]\frac{k_B T}{m\omega_o^{d_s}}\bar{A}^{\nu}. \tag{35}$$

Using $f(0) = 1$ and performing the integration in Eq. (27), we find (1) $\Phi[0] = \frac{6d_s\Gamma[\frac{d_s}{2}]}{2-d_s}$ in the Rouse model, and (2) $\Phi[0] = \frac{6d_s\Gamma[\frac{1}{d_w-d_f+1}]}{2-d_s}$ in the Zimm model ($\Gamma[x]$ is the Gamma function). Importantly, we have verified Eq. (31) for about 500 proteins by calculating its RW counterpart, the probability of return to the origin, which is exactly proportional to the time derivative of the Rouse model vibrational MSD [8].

The anomalous subdiffusion is expected to hold for times $\tau_0 \ll t \ll \tau_N$. For $t \gtrsim \tau_N$ the MSD saturates at $\langle \Delta\vec{u}(\infty)^2\rangle = 2\langle\vec{u}^2\rangle$; i.e., it is proportional to the so-called (mean) B-factors. As previously shown, the latter exhibits the generalized Landau-Peierls instability $\langle\vec{u}^2\rangle \sim N^{2/d_s-1}$. The Landau-Peierls instability can be deduced from the integral expression

$$\langle\vec{u}^2\rangle = 6d_s\frac{k_B T}{m\omega_o^{d_s}}\int_{\omega_{\min}}^{\omega_o} d\omega\, \omega^{d_s-3} \tag{36}$$

that for $d_s < 2$ diverges at the lower limit $\omega_{\min}$,

$$\begin{aligned}\langle\vec{u}^2\rangle &\simeq \frac{6d_s}{2-d_s}\frac{k_B T}{m\omega_o^{d_s}}\omega_{\min}^{d_s-2}\\ &\simeq \frac{6d_s}{2-d_s}\frac{k_B T}{m\omega_o^2}\left(\frac{R_g}{b}\right)^{d_w-d_f} \sim N^{2/d_s-1}.\end{aligned} \tag{37}$$

Interestingly, for $\tau_0 \ll t$ the MSD may be shown to obey a simple scaling formula that includes the crossover to saturation

$$\langle\Delta\vec{u}(t)^2\rangle = \langle\vec{u}^2\rangle\phi[t/\tau_N]. \tag{38}$$

Note that the anomalous diffusion exponent $\nu = (2 - d_s)/\theta$ can be deduced solely from the demand that, for $t \ll \tau_N$, the MSD should not depend on N. Thus $\phi[z] \sim z^{(2-d_s)/\theta}$ for $z \ll 1$, and $\phi[z] \to 2$ for $z \to \infty$.

E. Autocorrelation of distance

A closely related correlation function is the autocorrelation function $\langle\vec{x}(t)\cdot\vec{x}(0)\rangle$ of the fluctuations $\vec{x}(t)$ in the vector of separation between two points, where $\vec{x}(t)$ can be related to the displacements by $\vec{x}(t) = \vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',t)$. Note that $\langle\vec{x}(t)\cdot\vec{x}(0)\rangle$ is *not* identical to $\langle[\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2\rangle$. However, under disorder averaging we can exactly write the two-particle correlation function in question as the sum of the single-particle MSD and the distance autocorrelation function (see Appendix A 1),

$$\langle[\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2\rangle = \langle\Delta\vec{u}(t)^2\rangle + \langle\vec{x}(t)\cdot\vec{x}(0)\rangle. \tag{39}$$

(A similar expression holds without disorder averaging; see Appendix A 2.)

$\langle\vec{x}(t)\cdot\vec{x}(0)\rangle$ has been previously analyzed in great detail for short and long times. In particular, for $d_s < 2$ the static variance $\langle\vec{x}^{\,2}\rangle$ has been shown to diverge with distance as [5,8,38] $\langle\vec{x}^{\,2}\rangle \sim R_{ll'}^{d_w-d_f}$, where

$$R_{ll'} \equiv |\vec{R}_\ell - \vec{R}_{\ell'}| \tag{40}$$

is the Euclidean distance between beads $\vec{\ell}$ and $\vec{\ell}'$. This demonstrates the divergence with distance of the interparticle separation fluctuations (in accord with the Landau-Peierls instability), as recently verified numerically for about 500 proteins [8]. More precisely, putting $t = 0$ in Eqs. (39) and (23) we find, provided that $d_s < 2$ and $2\frac{d_s}{d_l} + d_s > 2$ [assuming $f(y) \simeq 1 - C_0 y^2$ for $y \ll 1$],

$$\begin{aligned}\langle\vec{x}^{\,2}\rangle &= 6d_s\frac{k_B T}{m\omega_o^{d_s}}\int_0^\infty d\omega\, \omega^{d_s-3}\{1 - f[(\omega/\omega_o)^{d_s/d_l}l]\}\\ &\simeq C\,\frac{k_B T}{m\omega_o^2}\left(\frac{R_{ll'}}{b}\right)^{d_w-d_f},\end{aligned} \tag{41}$$

where $C = 6d_s\int_0^\infty dz\, z^{d_s-3}[1 - f(z^{d_s/d_l})]$. Note that the static variance can be described by an effective harmonic potential $\frac{1}{2}m\omega_{\rm eff}^2\vec{x}^{\,2}$ with [21] $\omega_{\rm eff}^2 \approx \omega_o^2(R_{ll'}/b)^{d_f-d_w}$. Importantly, it has been proven proportional to the RW mean first passage time between $\vec{\ell}$ and $\vec{\ell}'$ on the same network, a fact that has far-reaching consequences [8,38,39]. For $d_s \geqslant 2$ a different behavior is obtained that is rarely relevant to proteins and is therefore not discussed here [8,38,39].

We repeat briefly the asymptotic analysis of $\langle\vec{x}(t)\cdot\vec{x}(0)\rangle$ for short and long times. From Eqs. (39) and (23) we find

$$\begin{aligned}\langle\vec{x}(t)\cdot\vec{x}(0)\rangle = 6d_s\frac{k_B T}{m\omega_o^{d_s}}&\int_{\omega_{\min}}^{\omega_o} d\omega\, \omega^{d_s-3}\\ &\times(1 - \bar{\Psi}(\omega_\alpha,|\vec{\ell} - \vec{\ell}'|))\mathrm{e}^{-\bar{A}\omega^\theta t},\end{aligned} \tag{42}$$

which may be expressed in the following scaling form:

$$\langle\vec{x}(t)\cdot\vec{x}(0)\rangle = \frac{k_B T}{m\omega_o^{d_s}}(\bar{A}t)^{\nu}\,\Omega[|\vec{\ell} - \vec{\ell}'|/\ell(t)], \tag{43}$$

where

$$\Omega[v] = 6d_s\int_0^\infty dz z^{d_s-3}(1 - f[z^{d_s/d_l}v])\mathrm{e}^{-z^\theta}. \tag{44}$$

Note the subtle, yet fundamental, difference between Eqs. (43)–(44) and Eqs. (26)–(27).

If $\ell(t) \ll |\vec{\ell} - \vec{\ell}'|$ (short times), the two beads'motion is uncorrelated, and (see Sec. IIID) each bead performs anomalous subdiffusion. To find the asymptote in this limit we add and subtract 1 to the integrand in Eq. (42) to obtain

$$\begin{aligned}\langle\vec{x}(t)\cdot\vec{x}(0)\rangle = 6d_s\frac{k_B T}{m\omega_o^{d_s}}&\int_{\omega_{\min}}^{\omega_o} d\omega\, \omega^{d_s-3}\{1 - \bar{\Psi}(\omega_\alpha,|\vec{\ell} - \vec{\ell}'|)\}\\ &- 6d_s\frac{k_B T}{m\omega_o^{d_s}}\int_{\omega_{\min}}^{\omega_o} d\omega\omega^{d_s-3}\{1-\bar{\Psi}(\omega_\alpha,|\vec{\ell}-\vec{\ell}'|)\}\\ &\times\left(1 - \mathrm{e}^{-\bar{A}\omega^\theta t}\right).\end{aligned} \tag{45}$$

We can then neglect $\bar{\Psi}(\omega_\alpha,|\vec{\ell} - \vec{\ell}'|)$ in the second integral in Eq. (45), consistent with $\ell(t) \ll |\vec{\ell} - \vec{\ell}'|$ and exponentially

decaying normal modes [see discussion after Eq. (10)]. This can be easily understood by changing the integration variable of the second integral to $z = \omega(\bar{A}t)^{1/\theta}$, and use $f[z^{d_s/d_l}|\vec{\ell} - \vec{\ell}'|/\ell(t)] \ll 1$ for nonvanishing z (since for vanishing z the contribution to the integral is negligible). The second integral then becomes identical to the MSD. Provided that $d_s < 2$, at short times we thus find

$$\langle \vec{x}(t) \cdot \vec{x}(0) \rangle \approx \langle \vec{x}^{\,2} \rangle - Bt^{\nu}, \tag{46}$$

where B is given by Eq. (35).

At long times such that $\ell(t) \gg |\vec{\ell} - \vec{\ell}'|$ the motion of the two particles is highly correlated, which leads to a vanishing autocorrelation of $\vec{x}(t)$. Using $f(y) \simeq 1 - C_0 \times y^2$ for $y \ll 1$ (C_0 is of order 1), we find, provided that $2 < 2\frac{d_s}{d_l} + d_s$,

$$\langle \vec{x}(t) \cdot \vec{x}(0) \rangle \simeq C_1 \frac{k_B T}{m\omega_o^{d_s(1+2/d_l)}} \left(\frac{R_{ll'}}{b}\right)^{2d_f/d_l} (\bar{A}t)^{-\mu}, \tag{47}$$

where $\mu = 2d_s/(d_l\theta) - \nu$, thereby

$$\mu = \frac{d_s}{d_l} + d_s/2 - 1 \tag{48}$$

in the Rouse model, and

$$\mu = \frac{2\frac{d_s}{d_l} + d_s - 2}{2 - d_s + d_s/d_f} = \frac{2\frac{d_f}{d_l} + d_f - d_w}{d_w - d_f + 1} \tag{49}$$

in the Zimm model. The numerical prefactor C_1 is $C_1 = 6d_s C_0 \Gamma[(d_s(1 + 2/d_l) - 2)/\theta]/\theta$.

To summarize the time dependencies, we find

$$\langle \vec{x}(t) \cdot \vec{x}(0) \rangle \sim \begin{cases} 1 - \text{const} \times t^{\nu} & \text{for} \quad t \ll t^*(R_{ll'}) \\ t^{-\mu} & \text{for} \quad t \gg t^*(R_{ll'}), \end{cases} \tag{50}$$

where

$$t^*(r) = \bar{A}^{-1}\omega_o^{-\theta}\left(\frac{r}{b}\right)^{d_f\theta/d_s}. \tag{51}$$

Thus $t^*(r) \sim r^{d_f\theta/d_s}$ and, in particular, $t^*(r) \sim r^{d_w}$ for the Rouse model.

F. Two-point correlation function: Scaling expressions

Combing the asymptotic results for $\langle \vec{x}(t) \cdot \vec{x}(0) \rangle$ with $\langle \Delta\vec{u}(t)^2 \rangle$ we obtain the following asymptotic behavior of the two-point correlation function under study. For short times, $t \ll t^*(R_{ll'})$, the correlation function is very close to its static value

$$\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle \simeq C \frac{k_B T}{m\omega_o^2} \left(\frac{R_{ll'}}{b}\right)^{d_w - d_f}. \tag{52}$$

Corrections to this result are of higher order than those discussed here. A crossover to a subdiffusive behavior occurs at $t \sim t^*(R_{ll'})$, and for long times, $t^*(R_{ll'}) \ll t \ll \tau_N$, we find

$$\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle \simeq B\, t^{\nu} + C_1 \frac{k_B T}{m\omega_o^{d_s(1+2/d_l)}} \left(\frac{R_{ll'}}{b}\right)^{2d_f/d_l} (\bar{A}t)^{-\mu}. \tag{53}$$

The second term in Eq. (53) is a small correction to the dominant time behavior $\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle \sim t^{\nu}$ that is independent of $R_{ll'}$.

To end this discussion, note that all scaling properties described in this section can be transformed from topological space to the real three-dimensional Euclidean space. Thus the two-point correlation function, Eq. (26), can be rewritten as

$$\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle = \frac{k_B T}{m\omega_o^{d_s}} (\bar{A}t)^{\nu}\ \Phi_1\left[\frac{R_{ll'}}{\xi(t)}\right], \tag{54}$$

where $\Phi_1[u] = \Phi[u^{d_f/d_l}]$. $\Phi_1[u]$ has the following asymptotes: (1) $\Phi_1[u] \simeq \text{const}$ for $u \ll 1$ (to leading order), or, more precisely, $\Phi_1[u] \simeq \Phi[0] + \text{const} \times u^{2d_f/d_l}$ (where the second term is negligible); and (2) $\Phi_1[u] \sim u^{d_w - d_f}$ for $u \gg 1$. Other equivalent expressions are

$$\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle = \frac{k_B T}{m\omega_o^{d_s}} (\bar{A}t)^{\nu}\ \Phi_2\left[\frac{t}{t^*(R_{ll'})}\right], \tag{55}$$

where $\Phi_2[u] = \Phi[u^{-d_s/(d_l\theta)}]$, and

$$\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle = \frac{k_B T}{m\omega_o^2} \left(\frac{R_{ll'}}{b}\right)^{d_w - d_f} \Phi_3\left[\frac{t}{t^*(R_{ll'})}\right], \tag{56}$$

where $\Phi_3[u] = u^{\nu}\Phi[u^{-d_s/(d_l\theta)}]$. $\Phi_3[u]$ has the following asymptotes: (1) $\Phi_3[u] \simeq \text{const}$ for $u \ll 1$, and (2) $\Phi_3[u] \sim u^{\nu}$ for $u \gg 1$ (to leading order), or, more precisely, $\Phi_3[u] \simeq \Phi[0]u^{\nu} + \text{const} \times u^{-\mu}$ (where the second term is negligible). Equation (56) is particularly useful for the numerical analysis that we perform next.

G. Numerical results for the Sierpinski gasket and proteins

1. *Sierpinski gasket*

To test the above analytic expression we first evaluate numerically the pair correlation function $\langle [\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2 \rangle$ on a vibrating Sierpinski gasket [18,20] [fractal and spectral dimensions: $d_f = \ln(3)/\ln(2) \simeq 1.585$, $d_s = 2\ln(3)/\ln(5) \simeq 1.3652$] having $N = 6561$ nodes (seventh generation) and gyration radius $R_g = 42.6$ nm. The values of model parameters were chosen such that they are similar to their values in native proteins: bond length and bead (amino acid) diameter $b = 5 \times 10^{-10}$ m, mass of a single bead $m = 4 \times 10^{-25}$ kg, spring natural frequency $\omega_o = 10^{12}\ \text{s}^{-1}$, viscosity of water $\eta = 8.94 \times 10^{-4}$ Pa s, and the temperature was set to $T = 298$ K. With these parameters the short cutoff, "amino-acid," relaxation time $\tau_0 = \gamma/\omega_o^2 = 3\pi\eta b/(m\omega_o^2) \simeq 10.53$ ps is within the range of typical values for proteins [23].

We solve numerically for the exact normal modes and eigen-frequencies of the Sierpinski gasket and evaluate the pair correlation function according to Eq. (21), in the framework of the Rouse model, for all pairs in the network that are separated by the same distance r within an interval $\delta r = 0.5$ Å. We then calculate the average value over all such pairs, for each point of time. We have used four values of r: 5, 15, 25, 50 Å. The results are shown in Fig. 1.

In Fig. 1(a) we plot, on a log-log scale, the four averaged two-point correlation functions versus the time t. Note the crossover from a constant value, which increases with increasing r as predicted by Eq. (52), to an anomalous subdiffusion time regime, identical to that of the single-particle MSD (effectively equal to the correlation function for $r = b = 5$A). In the subdiffusion regime, correlation functions obey Eq. (53),

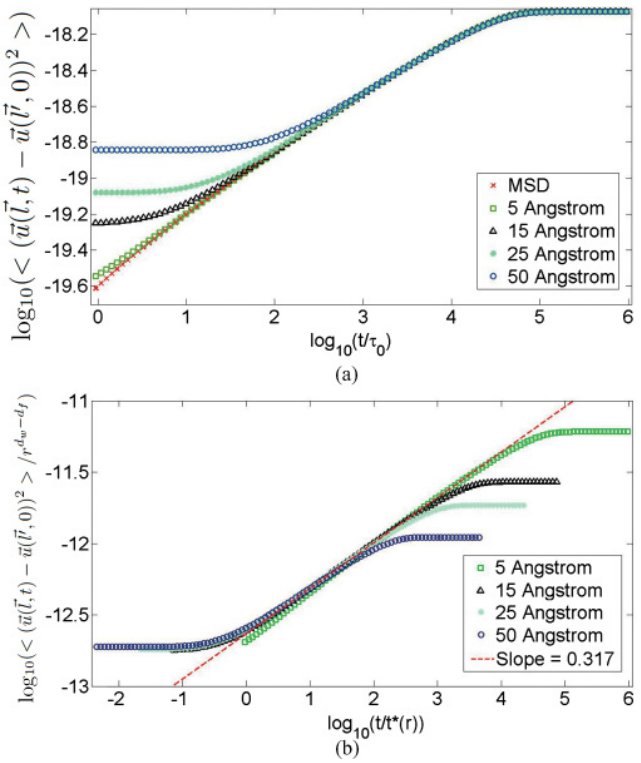


FIG. 1. (Color online) (a) Two-point correlation function for the Sierpinski gasket. The pair correlation function $\langle[\vec{u}(\vec{l},t)-\vec{u}(\vec{l}',0)]^2\rangle$ is evaluated numerically for bead pairs located on a vibrating Sierpinski gasket with 6561 nodes. Here, we focus on four groups of bead pairs, distanced $r = 5,\ 15,\ 25,\ 50 \pm \frac{1}{2}$ Å apart correspondingly. For a *fixed* value of the interbead distance r, $\langle[\vec{u}(\vec{l},t)-\vec{u}(\vec{l}',0)]^2\rangle$ is calculated for all pairs distanced $r \pm \frac{1}{2}$ Å apart. For every point in time, the correlation functions, in each distance group, are averaged over all pairs in that group. We plot the averaged two-point correlation functions vs the normalized time t/τ_0, on a log-log scale. Note the crossover from a constant value, which increases with increasing r as predicted by Eq. (52), to an anomalous subdiffusion time regime, identical to that of the single-particle MSD (plotted for comparison). As implied by Eq. (51), the crossover time to the anomalous subdiffusion regime increases with r. In the subdiffusion regime, the correlation function obeys Eq. (53). As the second term in Eq. (53) is only a small correction, behavior is expected to not depend on the interbead distance. Indeed, a common power-law behavior is observed for all distance groups. For long times, all curves saturate to a fixed value equal to $2\langle u^2\rangle_T$, i.e., twice the static MSD. (b) Scaled plot of (a). Following the predicted scaling behavior stated in Eq. (56), we normalize the correlation functions from (a) by $r^{d_w-d_f}$ and the time by $t^*(r) \sim r^{d_w}$. Data collapse to a single master curve is observed for $\tau_0 \ll t \ll \tau_N$. The slope ν in the subdiffusive time regime is found to be 0.317, in excellent agreement with the theoretical, Rouse model, value of $\nu = 1 - \frac{d_s}{2} \simeq 0.317$.

and the behavior is essentially independent of the distance r. Also note that the crossover time increases with increasing r as implied by Eq. (51). For long times, all curves saturate to the same value. This emerges from the fact that the saturation, $t \to \infty$, limit is always $2\langle\vec{u}^2\rangle_T$, and from the fact that in the Sierpinski gasket averaging over relatively small subset of nodes (i.e., those in the range of $r \pm \delta r$) is approximately identical to any other (e.g., different r) subset average, or to the complete spatial average.

In order to test the predicted scaling behavior of the two-point correlation function, in particular the version stated in Eq. (56), we normalize the correlation function by $r^{d_w-d_f}$ and the time by $t^*(r)$. The results are shown in Fig. 1(b). Data collapse to a single master curve is observed for times much longer than the shortest ("amino acid") vibrational relaxation time τ_0, and much shorter than the saturation time τ_N. Outside this range, data do not collapse to a single curve. Thus for $r = 5$A, the short time behavior does not obey scaling since for this distance $t^*(r) \simeq \tau_0$. Likewise, all curves diverge from each other close to saturation, $t \sim \tau_N$.

2. *Proteins*

Next, we perform numerical calculations of the pair correlation function for the protein LysX, PDB code 1UC8, containing $N = 505$ amino acids, using an identical procedure to the one done for the Sierpinski gasket. For this purpose the GNM is used with a cutoff distance $R_c = 6 \times 10^{-10}$ m and $m\omega_o^2$ is determined via fit of the mean theoretical (GNM) B-factor value to the experimental value reported in the PDB. Correcting for the over-stiffness of crystalline structures (studied by x rays) in comparison to structures in solution (studied by NMR), the value of $m\omega_o^2$ is further divided by a factor of four as suggested in Yang *et al.* [40] to yield $m\omega_o^2 = 0.1305$ N/m. The mass m is determined separately as the mean amino-acid mass and a value of $m = 1.66 \times 10^{-25}$ kg is obtained. The rest of the parameters are set as in the Sierpinski gasket: bead (amino acid) diameter $b = 5 \times 10^{-10}$ m, viscosity of water $\eta = 8.94 \times 10^{-4}$ Pa s, and temperature $T = 298$ K. Hence the molecular, "amino-acid," time is given by $\tau_0 = \gamma/\omega_o^2 = 32.29$ ps. The density of states $g(\omega)$ and mass distribution $M(r)$ of this protein are shown in Ref. [7], from which we have deduced the following values of the fractal and spectral dimension: $d_s = 1.73$, $d_f = 2.51$.

In Fig. 2(a) we plot, on a log-log scale, the averaged two-point correlation functions versus the time t, for a few separation distances r. Note again the crossover from a constant value, which increases with increasing r as predicted by Eq. (52), to an effective anomalous subdiffusion time regime, where the crossover time increases with increasing r as implied by Eq. (51). Here, unlike the Sierpinski gasket, curves do not saturate to the same value, showing that the average of $\langle\vec{u}^2\rangle_T$ over one given subset of amino acids (given r) may not be identical to the average over a different subset (different r), and may also differ from a full spacial average. Moreover, a clear subdiffusion regime is obtained only for the MSD ($r = 0$) due to the fact that τ_N is only a few times larger than $t^*(r)$ (finite-size effects) and the low value of the anomalous diffusion exponent ν.

In Fig. 2(b) we normalize the time by r^δ and the correlation function by r^α, and search for the exponents δ and α that collapse the data into a single master curve in the time regime $\tau_0 \ll t \ll \tau_N$. Unlike for the Sierpinski gasket, we avoid using the predicted exponents $\delta = d_w$ (for the Rouse model where $\theta = 2$) and $\alpha = d_w - d_f$, which are supposed to collapse the data. This is because, as found earlier [8], finite-size effects lead to modified exponents. Thus our scaling hypothesis is

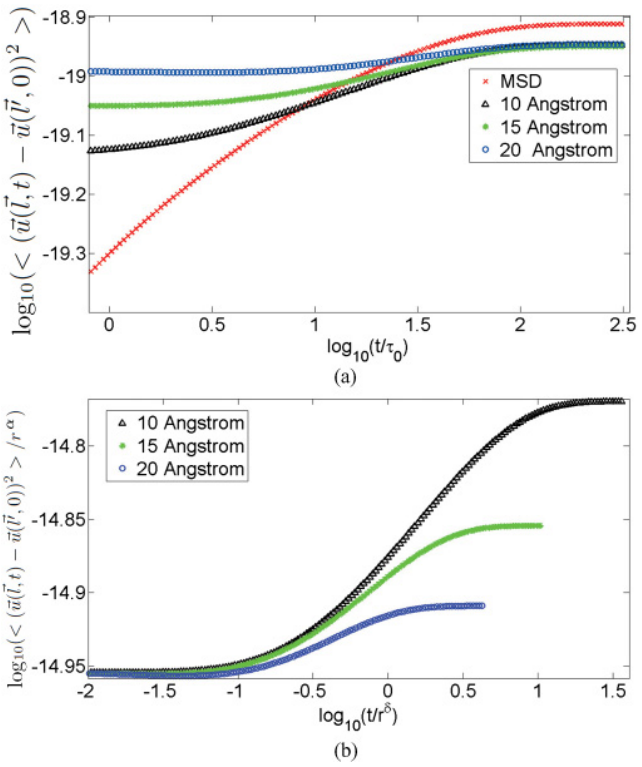


FIG. 2. (Color online) (a) Two-point correlation function for the protein LysX, PDB code 1UC8, $N = 505$. The pair correlation function $\langle[\vec{u}(\vec{l},t) - \vec{u}(\vec{l}',0)]^2\rangle$ is evaluated numerically for amino acids pairs in a procedure identical to the one used for the Sierpinski gasket. We plot, on a log-log scale, the averaged two-point correlation functions vs the the normalized time t/τ_0, for $r = 10,\ 15,\ 20$ Å. Note again the crossover from a constant value, that increases with r as predicted by Eq. (52), to an effective anomalous subdiffusion time regime. As implied by Eq. (51), the crossover time increases with r. Unlike the Sierpinski gasket, finite-size effects are clearly discernible. Curves do not saturate to the exact same value, showing that the average of $\langle u^2\rangle_T$ over a given subset of amino acids may not be identical to the average over a different subset and may also differ from the complete spacial average. In addition, a clear subdiffusion regime is apparent only for the MSD. (b) Scaled plot of (a). Similar to Fig. 1(a) we normalize the correlation functions from (a) by r^α and the time by r^δ. As finite-size effects may modify exponents [8], we allow α and δ to deviate from $d_w - d_f$ and d_w, respectively, and search for exponents that collapse the data into a single master curve in the time regime $\tau_0 \ll t \ll \tau_N$. We find that the values $\alpha = 0.464$ and $\delta = 3.09$ best collapse the data; these values are not too far from the predicted values $\alpha = 0.392$ and $\delta = 2.9$.

similar to Eq. (56) but with α replacing $d_w - d_f$ and δ replacing d_w [in $t^*(r)$]. We find that the values $\alpha = 0.464$ and $\delta = 3.09$ best collapse the data, and these values are not too far from the predicted values $\alpha = d_w - d_f = 0.392$ and $\delta = 2.90$ [7]. Thus we can infer the anomalous diffusion exponent for a protein whose topology is identical to LysX but whose size is "infinite," by demanding that for $t \gg t^*(r)$ the dependence on r disappears (ignoring the vanishing algebraically decaying term $\sim r^{2d_f/d_l} t^{-\mu}$). This implies $\Phi_3[u] \sim u^{\alpha/\delta}$ for $u \gg 1$ leading to $\nu = \alpha/\delta = 0.150$, quite close to the predicted exponent $\nu = 1 - d_s/2 = 0.135$.

IV. DYNAMIC STRUCTURE FACTOR

A. Introduction

The dynamic structure factor of complex fluids can be measured by various techniques, e.g, dynamic light scattering [41], inelastic neutron scattering [30], neutron spin-echo [25–27,42], and x-ray photon correlation spectroscopy [43]. All methods measure dynamic correlations of density fluctuations of the scatterers, and each method focuses on a different wavelength range and time regime. Among the four, the two relevant methods for detection of intramolecular protein dynamics are inelastic neutron scattering and neutron spin-echo. The individual scatterers will be defined here as the individual amino acids, and it will be assumed (as a simple but common approximation) that they all contribute equally to the scattering, irrelevant of their specific identity. The small variation in the scattering properties between one amino acid and the other is supposedly averaged out in the measurement. The structure factor is thus defined in the following way [41]:

$$S(\vec{k},t) = \frac{1}{V}\int_V d^3\vec{r}\int_V d^3\vec{r}' e^{i\vec{k}\cdot(\vec{r}-\vec{r}')}\langle c(\vec{r},t)c(\vec{r}',0)\rangle, \qquad (57)$$

where $c(\vec{r},t)$ is the stochastic number density, $\vec{k}$ is the scattering wave vector, and V is the macroscopic system volume.

To proceed, one inserts the stochastic number density [41]

$$c(\vec{r},t) = \sum_i \delta[\vec{r} - \vec{r}_i(t)], \qquad (58)$$

where the sum runs over *all* amino acids in the system, to obtain

$$S(\vec{k},t) = \frac{1}{V}\sum_{ij}\langle e^{i\vec{k}\cdot[\vec{r}_i(t)-\vec{r}_j(0)]}\rangle. \qquad (59)$$

We assume that the system is a dilute solution of N_p individual proteins $\{p\}$, all equivalent, each one having a different, yet random, orientation. Let $\vec{R}_p$ be the center-of-mass vector of protein p. We therefore have $\vec{r}_i(t) = \vec{R}_p + \vec{R}_{i,p}(t)$ where $\vec{R}_{i,p}$ is the position vector of amino acid i (represented by its α-carbon) in the center-of-mass coordinate frame of protein p, leading to

$$S(\vec{k},t) = \frac{1}{V}\sum_{p,p'=1}^{N_p}\langle e^{i\vec{k}\cdot[\vec{R}_p(t)-\vec{R}_{p'}(0)]}\rangle\sum_{ij=1}^{N}\langle e^{i\vec{k}\cdot[\vec{R}_{i,p}(t)-\vec{R}_{j,p'}(0)]}\rangle. \qquad (60)$$

For dilute solutions, the internal dynamics in each protein is decoupled from the internal dynamics of all other proteins, yet, as they are all identical, it bears exactly the same stochastic evolution. In addition, proteins diffuse independently of each other. Thus, the summation over terms with $p \neq p'$ contributes only to $\vec{k} = 0$ ("forward scattering"), and for $\vec{k} \neq 0$ only terms with $p = p'$ survive to give [41]

$$S(k,t) = \frac{N}{V}\sum_{p=1}^{N_p}\langle e^{i\vec{k}\cdot[\vec{R}_p(t)-\vec{R}_p(0)]}\rangle\langle S_p(\vec{k},t)\rangle_{\vec{k}}, \qquad (61)$$

where N is the number of amino acids in a single protein. Here we have defined the single-protein structure factor (with $1/N$ normalization factor)

$$S_p(\vec{k},t) = \frac{1}{N}\left\langle \sum_{i,j=1}^{N} e^{i\vec{k}\cdot[\vec{R}_i(t)-\vec{R}_j(0)]} \right\rangle, \tag{62}$$

where $\vec{R}_i$ is the position vector of amino acid i in the center of mass coordinate frame and the sum runs over all amino acids in a single protein. $S_p(k,t) \equiv \langle S_p(\vec{k},t)\rangle_{\vec{k}}$ denotes angular averaging, i.e., averaging over all different orientations of the protein with respect to the scattering wave vector $\vec{k}$. This is required since each protein participating as a scattering center is, at any particular time, at random orientation with respect to $\vec{k}$.

If proteins perform simple diffusion in the solvent we find [33,41]

$$S(k,t) = c\ e^{-k^2 D_{\rm cm} t} S_p(k,t), \tag{63}$$

where c is the (mean) amino acid number density, i.e., $c = Nc_p$ where $c_p = N_p/V$ is the protein number density, and $D_{\rm cm}$ is the center of mass diffusion coefficient of a single protein. The latter can be estimated using the Stokes-Einstein relation, $D_{\rm cm} \simeq k_B T/6\pi\eta R_h$, where R_h is the hydrodynamic radius, which is approximately proportional to the gyration radius R_g, with a proportionality constant that weakly depends on the number of beads (amino acids) N and fractal dimension d_f [33,44,45]. Rotational diffusion [46] is considered explicitly in the subsequent section and in Appendix B as part of the (ensemble-averaged) protein structure factor.

B. Protein structure factor: Exact results

We now calculate the ensemble average structure factor of proteins in dilute solution. First, consider a protein having a fixed but arbitrary orientation with respect to the wave vector $\vec{k}$, with rotational motion arrested. Orientational average will be performed at a later stage. Writing $\vec{R}_i(t) = \vec{R}_{i,\rm eq} + \vec{u}_i(t)$ we obtain

$$S_p(\vec{k},t) = \frac{1}{N}\sum_{i,j=1}^{N} e^{i\vec{k}\cdot(\vec{R}_{i,\rm eq}-\vec{R}_{j,\rm eq})}\left\langle e^{i\vec{k}\cdot[\vec{u}_i(t)-\vec{u}_j(0)]}\right\rangle. \tag{64}$$

Note that, relating to the notations of Secs. II and III, $\vec{u}_i \equiv \vec{u}(\vec{l})$ and $\vec{u}_j \equiv \vec{u}(\vec{l}')$.

Consider now the stochastic variable $\vec{u}_i(t) - \vec{u}_j(0)$. Recall that the dynamics of $\vec{u}_i(t)$ is governed by the linear Langevin equation (11), in which the stochastic force is a white noise obeying Gaussian statistics. By a general theorem of stochastic processes [47], the statistics of $\vec{u}_i(t)$, being the solution of Eq. (11), is also Gaussian, and so is the combination $\vec{u}_i(t) - \vec{u}_j(0)$. Using the well-known property of Gaussian fluctuations [33,47], and the isotropy of scalar elasticity (GNM), i.e., the fact that the statistics of all three components of $\vec{u}_i(t)$ is identical, we thus find

$$\langle e^{i\vec{k}\cdot[\vec{u}_i(t)-\vec{u}_j(0)]}\rangle = e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}. \tag{65}$$

Using Eqs. (65) in (64) we have, omitting henceforth the subscript "eq" in $\vec{R}_{i,\rm eq}$,

$$S_p(\vec{k},t) = \frac{1}{N}\sum_{i,j=1}^{N} e^{i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)} e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}. \tag{66}$$

Performing angular averaging we find

$$S_p(k,t) = \frac{1}{N}\sum_{i,j=1}^{N} \frac{\sin[kR_{ij}]}{kR_{ij}} e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}, \tag{67}$$

where $R_{ij} = |\vec{R}_i - \vec{R}_j|$ is the Euclidean, real-space, distance between amino acids (represented by their α-carbons) i and j. Equation (67) is a good starting point for numerical evaluation of both the static ($t = 0$) and dynamic protein structure factor for specific proteins based on their three-dimensional structure. Noteworthy is that this expression does not include any unfolding-refolding dynamics which may occur on the submicrosecond time scale or longer.

When short time rotational diffusion is included, followed by the required angular averaging, the result is (see Appendix B)

$$S_p(k,t) \simeq \frac{1}{N}\sum_{i,j=1}^{N} \left\langle e^{i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)} e^{-k^2 R_i^2 \sin^2\theta_i D_{\rm rot} t}\right\rangle_{\vec{k}} \times e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}, \tag{68}$$

where $D_{\rm rot}$ is the rotational diffusion coefficient ($D_{\rm rot} t \ll 1$ is assumed), θ_i is the angle between $\vec{k}$ and R_i, and $\langle\cdots\rangle_{\vec{k}}$ means angular average over all directions of $\vec{k}$ at fixed protein orientation. The rotational diffusion coefficient of a protein can be estimated from the Stokes-Einstein-Debye expression [48] $D_{\rm rot} \simeq k_B T/8\pi\eta R_g^3$. The angular averaging in Eq. (68) requires numerical integration for accurate evaluation. In the following section and Appendix B it is also evaluated approximately to obtain the dominant behavior.

C. Protein structure factor: Approximate results

In this section we present approximate and scaling expressions that can be obtained for fractals and proteins (based on their fractal-like structure). In doing so, we rely on results for the pair correlation function discussed in Sec. III, $\langle[\vec{u}(\vec{\ell},t) - \vec{u}(\vec{\ell}',0)]^2\rangle$ [recall that $\vec{u}_i \equiv \vec{u}(\vec{l})$]. The details of the approximations are given in Appendix C.

Consider first the protein static structure factor. To examine the effect of vibrations on the static structure factor we define a "frozen network" structure factor in which vibrations are arrested:

$$\tilde{S}_p(k) = \frac{1}{N}\sum_{i,j} \frac{\sin[kR_{ij}]}{kR_{ij}}. \tag{69}$$

For $kR_g \gg 1$ we obtain the well known result $\tilde{S}_p(k) \sim k^{-d_f}$, which can be also obtained by scaling approach [19,33,49,50]. When vibrations are included we find, to leading order, the first-order correction to $S_p(k) \sim k^{-d_f}$ in the form

$$S_p(k) \approx \tilde{S}_p(k) - \text{const} \times\ k^{2-d_w} + \cdots. \tag{70}$$

However, due to the smallness of the prefactor of the second term, this correction is negligible and $S_p(k) \simeq \tilde{S}_p(k)$.

Next, we consider the protein dynamic structure factor. It will be later shown that the rotational diffusion time dependence in $S_p(k,t)$ is rather weak for $t \lesssim \tau_N$. We shall therefore evaluate first the dynamic structure factor with rotational diffusion arrested Eq. (67). For short times, $t \ll \tau(k)$, where $\tau(k) \sim k^{-d_f\theta/d_s}$, information did not have time to negotiate a "blob" of linear size $\sim k^{-1}$, and we find that the dynamic structure factor did not decay much and is left almost equal to the static structure factor, $S_p(k,t) \simeq S_p(k)$. A more accurate description yields, to leading order,

$$S_p(k,t) \simeq S_p(k) - \text{const} \times \; k^2\, t^{2/\theta}. \tag{71}$$

For the Rouse model ($\theta = 2$) this implies that the short time decay is roughly exponential.

At longer times, $\tau(k) \ll t \ll \tau_N$, i.e., when $1 \ll k\xi(t) \ll kR_g$, information has propagated beyond the scattering wavelength $\sim 1/k$. Physically, this implies that the "blob" of size $\sim 1/k$, which is controlling the relaxation at wave vector $\vec{k}$, is now moving almost coherently as if it was a single bead. This leads to a *stretched exponential* decay of the dynamic structure factor,

$$S_p(k,t) \simeq \tilde{S}_p(k)\, \exp[-(\Gamma_k t)^{\nu}], \tag{72}$$

where

$$\Gamma_k = (B/6)^{1/\nu} k^{2/\nu}. \tag{73}$$

Note that the stretching exponent is exactly the anomalous diffusion exponent ν. The stretched exponential decay, together with the dependence of the stretching exponent ν, is thus a strong signature of the fractal structure.

Next we estimate the effect of rotational diffusion using Eq. (68) at short times, $D_{\rm rot} t \ll 1$. We find that the effect is highly sensitive to the value of k and the observed time regime. Two regimes of time can be distinguished depending on whether $kR_g\sqrt{D_{\rm rot} t}$ is smaller or larger than unity. If $kR_g\sqrt{D_{\rm rot} t} \gg 1$, as can occur for very large k and not too short times, we find in Appendix C that the contribution of rotational diffusion leads, to leading order, to a power law, $\sim t^{-d_f/2}$, pre-exponential factor. Adding translational diffusion according to Eq. (63), we find

$$S(k,t) \approx \text{const} \times c\; \tilde{S}_p(k)\, t^{-d_f/2}\, \exp[-(\Gamma_k t)^{\nu}] \times \exp\left[- k^2 \frac{k_B T}{6\pi\eta R_g} t \right]. \tag{74}$$

In the opposite limit where $kR_g\sqrt{D_{\rm rot} t} \ll 1$ (but still large wave numbers $kR_g \gg 1$), we obtain for the combined effect of rotational-translational diffusion

$$S(k,t) \approx c\; \tilde{S}_p(k)\, \exp[-(\Gamma_k t)^{\nu}] \times \exp\left[- k^2 \frac{3d_f + 4}{12\pi(d_f+2)} \frac{k_B T}{\eta R_g} t \right]. \tag{75}$$

In case that $R_g \gg b$ (not quite the situation for proteins) we have $Bt^{\nu} \gg (k_B T/\eta R_g)t$ in the relevant time regime $t \ll \tau_N$. In this limit Eq. (75) reduces to a pure stretched exponential decay, $S(k,t) \approx c\; \tilde{S}_p(k)\, \exp[-(\Gamma_k t)^{\nu}]$.

D. Numerical results for the Sierpinski gasket and proteins

1. *Sierpinski gasket*

As done for the pair correlation function, we first evaluate numerically the single-particle dynamic structure factor $S_p(k,t)$ for the Sierpinski gasket. Using the single-molecule structure factor Eq. (67), where rotational and translational diffusion are arrested, allows us to focus on the intramolecular vibrations. As argued above, for $kR_g \gg 1$ rotational and translational diffusion have negligible effect in the measurable time regime where the structure factor did not yet decay to a vanishing value. For simplicity and clarity, we use below only the Rouse model of friction, although our conclusions apply equally to the Zimm model. We calculate the dynamic structure factor for four generations of the Sierpinski gasket such that the number of nodes N in the gasket (i.e., gasket "size") varies between $N \simeq 200$ and $N \simeq 7000$.

In Fig. 3 we plot the static and dynamic structure factors for the Sierpinski gasket. The inset shows the static structure factor

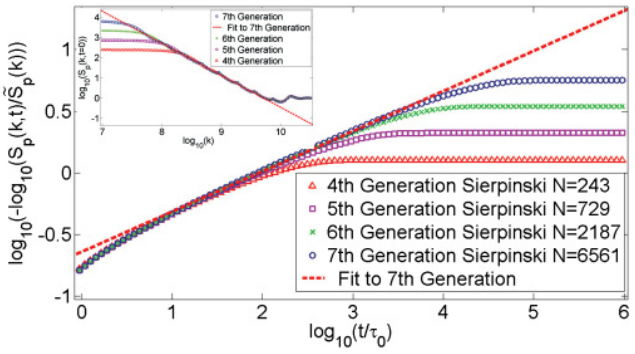


FIG. 3. (Color online) Dynamic structure factor of Sierpinski gaskets of various sizes. In order to demonstrate the stretched exponential decay of the dynamic structure factor for fractal objects we plot $\log_{10}[-\log_{10}[S(k,t)/\tilde{S}(k)]]$ vs $\log_{10}(t/\tau_0)$ for Sierpinski gaskets of various sizes. Here $\tilde{S}(k)$ is the frozen structure factor, $k = 10^{10}$ m^{-1}, $\tau_0 = 10.53$ ps, and t ranges from 10 ps and up to 10 μs. A clear stretched exponential decay, the onset of which starts at $t \simeq 10\tau_0$, is visible for all gaskets. As expected, the time regime into which this decay extends is shown to grow with the size of the gasket. For the largest gasket (seventh generation, $N = 6561$, $R_g = 42.6$ nm), the stretched exponential decay persists for approximately three decades. Fitting the data for this gasket, in the time interval $\log_{10}(t/\tau_0) \in (0.81, 3.9)$, we find an exponent of $\simeq 0.325$, with excellent agreement with the theoretical value of $1 - \frac{d_s}{2} \simeq 0.317$. Inset: The static structure factor of the Sierpinski gasket. We plot $S(k, t = 0)$ vs the wave number k on a log-log scale. Here k ranges from 10^7 m^{-1} and up to 3.16×10^{10} m^{-1}. A clear power-law decay, terminating at roughly $k \simeq 10^{9.5}$ m^{-1}, is visible for all gaskets. As expected, the wave-number regime in which this decay is observed is shown to grow with the size of the gasket. For the largest gasket the power-law decay persists for about two decades. Fitting the data for this gasket, in the wave-number interval $\log_{10}(k) \in (7.8, 9.5)$, we find an exponent of -1.583. This value stands in excellent agreement with the theoretical value of $d_f \simeq 1.585$. We note that in contrast to the frozen structure factor $\tilde{S}(k)$, the static structure factor $S(k, t = 0)$ does take into account the contribution of thermal vibrations (see main text). However, as is evident from the plot, vibrations have a negligible effect on the static structure factor for realistic spring constants and temperatures.

on a log-log scale for the largest gasket in the series. Note the clear $S_p(k) \sim k^{-d_f}$ power-law behavior in the regime $R_g^{-1} \ll k \ll b^{-1}$, showing that vibrations have negligible effect on the static structure factor for realistic spring constants and temperatures. The dynamic structure factor (main figure) of the four gasket generations is shown as $-\log_{10}[S_p(k,t)/\tilde{S}_p(k)]$ versus t/τ_0 on a log-log scale, such that a stretched exponential decay would show on this plot as a straight line whose (positive) slope is the stretching exponent. Note that we divided the dynamic structure factor by the "frozen network" static structure factor, $\tilde{S}_p(k)$, in accord with Eq. (72). The results are presented for a particular wave number, $k = 10^{10}$ m^{-1} for which $kR_g \simeq 426 \gg 1$ for the largest gasket in the series ($N = 6561$), and $kR_g \simeq 53 \gg 1$ for the smallest gasket ($N = 243$). Note that a straight line, whose slope is positive, is formed on an intermediate time window that widens up as we move from a smaller to a larger gasket, demonstrating the diminishing contribution of finite-size effects as the system size increases. The stretching exponent that is obtained from the plot, $\beta \simeq 0.325$, is very close to the theoretical value $\beta = \nu = 1 - d_s/2 \simeq 0.317$.

In Fig. 4(a) we plot the dynamic structure factor for the largest gasket studied ($N = 6561$) for different wave numbers k, plotting $-\log_{10}[S_p(k,t)/S_p(k,t=0)]$ versus t/τ_0 on a log-log scale. Note that here we divided the dynamic structure factor by the true static structure factor $S_p(k) \equiv S_p(k,t=0)$. This is done in order to verity the quality of the stretched exponential behavior when such an experimental-type analysis of the data is being performed, since $\tilde{S}_p(k)$ is not a measurable quantity, unlike $S_p(k)$. A clear straight line whose slope is positive is formed on an intermediate time window that widens up as k increases. This is in accord with the theoretical prediction associating the stretched exponential behavior with the time regime $\tau(k) \ll t \ll \tau_N$ where $\tau(k) \sim k^{-d_f\theta/d_s} \sim k^{-2.32}$. For the smaller k studied one can also observe an early exponential-like decay, shown as a straight line with slope $\simeq 1$ on this plot, in accord with Eq. (71). This exponential decay slowly crosses over to the stretched exponential decay. Despite the use of $S_p(k)$ rather than $\tilde{S}_p(k)$, the numerically obtained stretching exponent for the largest k in the series (upper curve), $\beta \simeq 0.30$, is remarkably close the theoretical value $\beta = \nu = 1 - d_s/2 \simeq 0.317$, demonstrating that this experimental-like analysis yields reliable exponents. Fig. 4(a) is accompanied by Fig. 4(b) in which we perform local slope analysis. This is done by plotting the first derivative of $\log_{10}[-\log_{10}[S_p(k,t)/S_p(k)]]$ versus $\log_{10}[t/\tau_0]$. Note the very clear "shoulders" whose width increase for increasing k, signifying constant slopes in Fig. 4(a). Figure 4(b) thus provides a complementary assessment of the stretched exponential behavior.

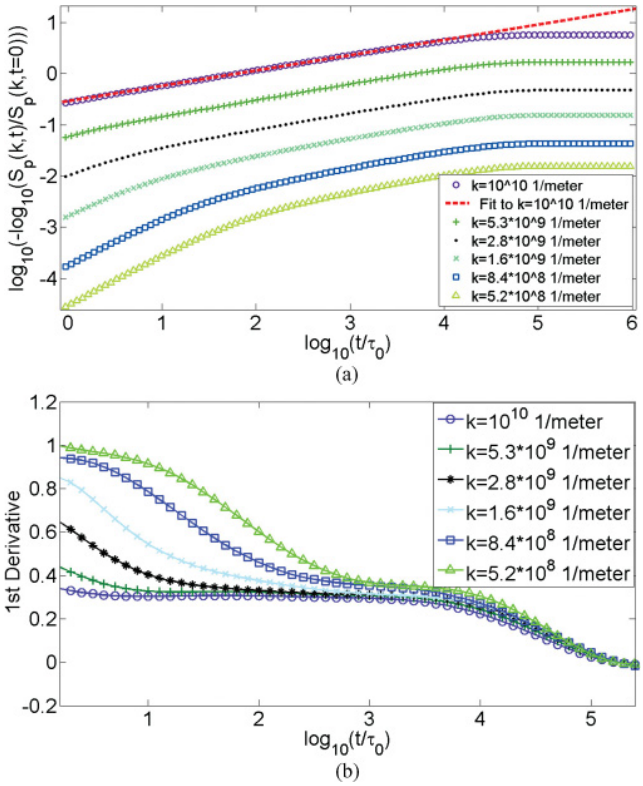


FIG. 4. (Color online) (a) Dynamic structure factor of the Sierpinski gasket-wave-number dependence. In order to demonstrate the dependence of the stretched exponential decay upon the wave number k we plot $\log_{10}[-\log_{10}[S(k,t)/S(k,t=0)]]$ vs $\log_{10}(t/\tau_0)$ for various values of k. Here $S(k,t=0)$ is the static structure factor, $\tau_0 = 10.53$ ps, and t ranges from 10 ps up to 10 μs. All data are plotted for the seventh-generation Sierpinski gasket ($N = 6561$, $R_g = 42.6$ nm). As expected, as the wave number k decreases, the onset of the stretched exponential decay [roughly proportional to $\tau(k) \sim k^{-d_f\theta/d_s}$] is pushed toward later times. For $k = 10^{10}$ m^{-1} we fit the data in the time interval $\log_{10}(t/\tau_0) \in (0.81,3.9)$ and find an exponent of 0.30, in excellent agreement with the theoretical value of $1 - \frac{d_s}{2} \simeq 0.317$. Note that in this figure we have used $S(k,t=0)$ as a normalizing factor for $S(k,t)$. This is done in order to verify the quality of the stretched exponential behavior when such an experimental-type analysis of the data is being performed, since $\tilde{S}(k)$ is not a measurable quantity, unlike $S(k)$. (b) Local slope analysis of the data plotted in (a). The first derivative of $\log_{10}[-\log_{10}[S(k,t)/S(k)]]$, with respect to $\log_{10}[t/\tau_0]$, is plotted vs $\log_{10}[t/\tau_0]$. A clear plateau ("shoulder") region is formed on an intermediate time window that widens up as k increases. For the smallest k studied, this region is visible for $\log_{10}[t/\tau_0] \in (3,3.75)$, and for the largest k studied it extends across $\log_{10}[t/\tau_0] \in (0.5,3.75)$. The height of the plateau region corresponds to the value of the exponent that characterizes the stretched exponential decay. As k increases, the height of the plateau region slowly converges to a value of $\simeq 0.30$, remarkably close the theoretical value $\beta = \nu = 1 - d_s/2 \simeq 0.317$.

2. *Proteins*

As done for the Sierpinski gasket, we perform numerical evaluation of the protein dynamic structure factor $S_p(k,t)$, Eq. (67), arresting the rotational and translational diffusion. We study three different proteins of variable sizes: 1FTR ($N = 1184$), 1UC8 ($N = 505$, studied in Sec. IIIG2), and 3TSS ($N = 190$). The GNM is used and parameters are chosen as described in Sec. IIIG2. Parameter values slightly differ form one protein to another, and we find $m\omega_o^2 =$ 0.1898, 0.1305, 0.2510 N/m and $m = 1.65 \times 10^{-25}$, 1.66×10^{-25}, 1.76×10^{-25} kg respectively. This leads to the following value of molecular, "amino-acid," time $\tau_0 = \gamma/\omega_o^2 =$ 22.19, 32.29, 16.78 ps, respectively. The density of states $g(\omega)$ and and mass distribution $M(r)$ of the three proteins

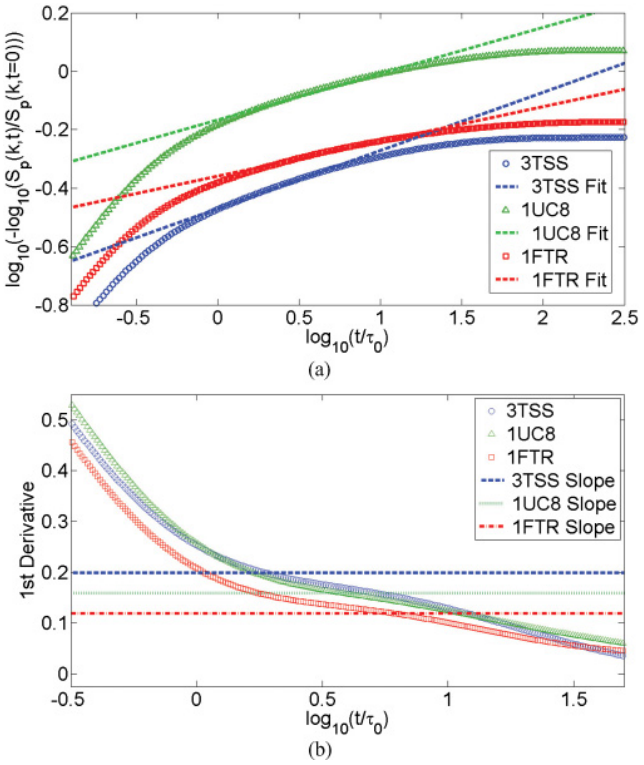


FIG. 5. (Color online) (a) Dynamic structure factor of proteins. $-\log_{10}[S_p(k,t)/S_p(k)]$ vs t/τ_0 is plotted on a log-log scale for three different proteins, 1FTR ($N = 1184$), 1UC8 ($N = 505$), and 3TSS ($N = 190$). Here $\tau_0 = \gamma/\omega_o^2 = 22.19,\ 32.29,\ 16.78$ ps, respectively, and $k = 10^{10}$ m^{-1}. As in Fig. 4(a), we divide the dynamic structure factor by the static structure factor $S_p(k) \equiv S(k,t=0)$. The decay is clearly nonexponential. However, unlike in Fig. 4(a) we do not observe a clear straight line in an intermediate time window that spans across several time decades, indicating that the decay is not a pure stretched exponential. The effective stretching exponents that are obtained in an approximate fit are $\beta = 0.119,\ 0.1584,\ 0.199$, respectively. (b) Local slope analysis of the data plotted in (a). The first derivative of $\log_{10}[-\log_{10}[S(k,t)/S(k)]]$, with respect to $\log_{10}[t/\tau_0]$, is plotted vs $\log_{10}[t/\tau_0]$. The horizontal dashed lines are the slope values obtained from the fits in (a). It can be seen that these lines cross the derivative lines in the middle of a weak "shoulder" that signifies an approximate stretched exponential regime.

is shown in Ref. [7], from which we take the following values of the fractal and spectral dimension: 1FTR–$d_s = 1.93$, $d_f = 2.66$; 1UC8–$d_s = 1.73$, $d_f = 2.51$; 3TSS–$d_s = 1.52$, $d_f = 2.50$. Since each protein has a different d_s value, we expect a different stretching exponent $\beta = \nu = 1 - d_s/2$ for each one: $\nu = 0.035,\ 0.135,\ 0.24$, respectively.

In Fig. 5(a) we plot the dynamic structure factor of the proteins mentioned above. The theoretical prediction associates a stretched exponential behavior in the time regime $\tau(k) \ll t \ll \tau_N$ that translates to $0.1\tau_0 \ll t \ll 70\tau_0$, and the fitted data shown in Fig. 5(a) are well within this regime. Indeed, within this intermediate regime the decay is clearly nonexponential, a fingerprint of the fractal-like structure. However, due to finite-size effects, the stretching exponents (i.e., the slopes) obtained from the fits, $\beta = 0.119,\ 0.1584,\ 0.199$, respectively, deviate from the predicted values of ν stated above. Figure 5(a) is accompanied by Fig. 5(b) in which we perform local slope analysis. This is done by plotting the first derivative of $\log_{10}[-\log_{10}[S_p(k,t)/S_p(k)]]$ versus $\log_{10}[t/\tau_0]$. It is possible to observe very weak shoulders, signifying the existence of roughly constant slopes, in the regimes that correspond to the fitted regimes in Fig. 5(a) (apparently more visible for 3TSS). It appears that finite-size effects are strong in proteins and that a pure stretched exponential behavior is hardly obtained when we consider vibrations alone. Although one could have naively expected (in analogy with the Sierpinski gasket) that for proteins with $\sim$1000 amino acids or more (e.g., 1FTR in this study) finite-size effects would be less significant, this is not the case. As it turns out, in such proteins the value of d_s is usually very close to 2, consequently leading to vanishing stretching exponents ν. Such an exponent is difficult to determine accurately as it is highly sensitive to the short and long time, non-stretched-exponential, dynamics. Nevertheless, the values obtained from our fits of the effective exponents quantify the nonexponential behavior as a whole. Since the GNM (and hence our numerical calculations) are solely based on protein structure, the nonexponential behavior we observe is a direct manifestation of the fractal-like structure of proteins.

3. *Proteins: Including rotational and translational diffusion and comparison with neutron spin-echo experiments*

Next we further examine the decay of the dynamic structure factor when rotational and translational diffusion are included. Consider, first, the effect of rotational diffusion for the above three proteins. Adding rotational diffusion according to Eq. (68) [or Eq. (B11)], using $R_h = R_g$ where R_g is calculated from the PDB structure, we plot in Fig. 6 $-\log_{10}[S_p(k,t)/S_p(k)]$ versus t/τ_0 on a log-log scale. We observe a clear, effective, stretched exponential decay of the dynamic structure factor in an intermediate time regime, yielding stretching exponents $\beta = 0.34,\ 0.28,\ 0.32$ for 3TSS, 1UC8, and 1FTR, respectively. Note that these exponents are higher than those obtained from vibrations alone. The inset in Fig. 6 includes also the contribution of translational diffusion using again $R_h = R_g$. Following Ref. [27], we define $t_0 = 0.01$ ns, and plot $S(k,t)/S(k,t_0)$ versus the time in nanoseconds. The cumulative effect of all three dynamical processes gives rise to a nonexponential decay that can be well fitted by a stretched exponential in the time interval $t \in [0.01\text{ ns}, 1\text{ ns}]$. The effective stretching exponents are $\beta = 0.72,\ 0.42,\ 0.65$ for 3TSS, 1UC8, and 1FTR, respectively, and are significantly higher than those obtained from vibrations and rotations alone. We conclude that, in proteins, the exact value of the effective stretching exponent is sensitive to the interplay between vibrations, rotations, and translations.

Next we consider two proteins, horse heart myoglobin (Mb, PDB code 3LR7) and bovine hemoglobin (Hb, PDB code 2QSP), that have been recently investigated in dilute solutions by neutron spin echo. Analyzing the vibrational density of states of these proteins we find $d_s = 1.56$ for Mb and $d_s = 1.74$ for Hb, leading to $\nu = 1 - d_s/2 = 0.22,\ 0.13$ for Mb and Hb, respectively. Adding translational and rotational diffusion, and normalizing the dynamic structure factor by the effective static structure factor at $t = 10$ ps (similar to the procedure in Ref. [27]), we find that for the experimental value of k, $k = 0.579\text{A}^{-1}$, an effective stretched exponential decay

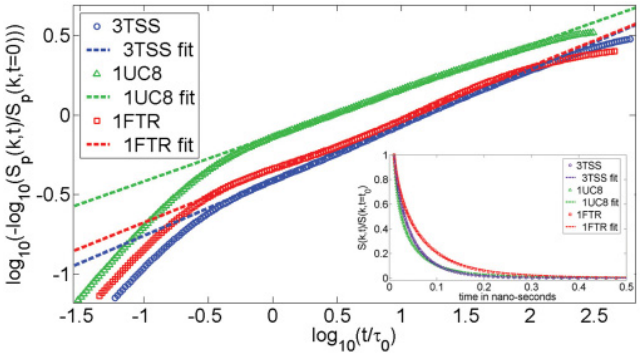


FIG. 6. (Color online) The joint effect of vibrations and rotations on the dynamic structure factor of proteins. Numerically evaluating $S_p(k,t)$, Eq. (68) [or Eq. (B11)], we study the joint effect of vibrations and rotations on the dynamic structure factor. We plot $-\log_{10}[S_p(k,t)/S_p(k)]$ vs t/τ_0 on a log-log scale for the same three proteins that appear in Fig. 5. We observe a clear, effective, stretched exponential decay of the dynamic structure factor in an intermediate time regime. The effective stretching exponents are $\beta = 0.34, 0.28, 0.32$ for 3TSS, 1UC8, and 1FTR, respectively. Note that these exponents are higher than the ones obtained when only vibrations were taken into account. Inset: The contribution of translations. Numerically evaluating $S(k,t)$, Eq. (63), we add the contribution of translations to that of vibrations and rotations. We now follow Ref. [27], define $t_0 = 0.01$ ns, and plot $S(k,t)/S(k,t_0)$ vs the time in nanoseconds. The cumulative effect of all three processes give rise to a nonexponential decay that can be well fitted by a stretched exponential in the time interval $t \in [0.01\text{ ns}, 1\text{ ns}]$. The effective stretching exponents are $\beta = 0.72, 0.42, 0.65$ for 3TSS, 1UC8, and 1FTR, respectively. Note again that, as expected, these exponents are higher than the ones obtained when only vibrations and rotations are taken into account. Since the decay due to translations is purely exponential, we conclude that the origins of the apparent stretched exponential are the contributions due to vibrations and rotations. However, the exact value of the effective stretching exponent is sensitive to the interplay between vibrations, rotations and translations. It is therefore nonuniversal and may vary from protein to protein and with external conditions.

is obtained in the experimental time window 10 ps $\leqslant t \leqslant$ 1 ns, with stretching exponents $\beta = 0.89$ (Mb) $\beta = 0.86$ (Hb). These exponents are quite close to unity, suggesting that in the case of these two proteins, either that effect the vibrations have been underestimated or that the translational-rotational effect has been overestimated. Support for the latter possibility appears in the data of Ref. [27] that shows that the translational diffusion coefficient is concentration dependent, suggesting that the solution is (perhaps) not sufficiently dilute to prevent aggregation (oligomerization) of proteins so that aggregate (oligomer) diffusion coefficients should be used instead of those of individual proteins. Another possibility that can lead to a large hydrodynamic radius is the formation of a thick hydration shell. Exploring this option we introduce an effective radius R_g^*, vary it between the single-protein value R_g and four times this value and use it to evaluate $D_{\rm rot}$ and $D_{\rm cm}$. We find that for $R_g^* \simeq 4R_g$, the stretching exponent lowers to $\beta \simeq 0.75$. Moreover, increasing slightly the value of k to $k = 1$ A^{-1} the stretching exponent further goes down to $\beta \simeq 0.52$. This shows that the exact value of the stretching exponent is not universal and results from an interplay between vibrations, rotations, and translations. However, it is gratifying that one can obtain a low value of β, due to the very low value of ν that serves as an effective lower bound for β.

V. CONCLUSIONS

We have integrated and advanced upon our previous studies on the vibrational dynamics of fractals in general, and of proteins as fractals, and presented a thorough study of the dynamic structure factor $S(k,t)$ of vibrating fractals and proteins in dilute solutions. Our study is highly detailed and may allow comprehensive comparison with experiment. Our main result is, however, simple and shows that so long as $kR_g \gg 1$ and $k\bar{u} \gtrsim 1$, where $\bar{u} \equiv \sqrt{\langle u^2\rangle}$, and with rotational and translational degrees of freedom arrested, the decay of the dynamic structure factor is strongly influenced by the anomalous diffusion of amino acids (beads) at short times, $\langle \Delta\vec{u}(t)^2\rangle = B\, t^{\nu}$. The value of ν depends on the fractal (d_f) and spectral (d_s) dimensions, $\nu = 1 - d_s/2$ in the Rouse model of friction and $\nu = (2-d_s)/(2-d_s+d_s/d_f)$ in the Zimm model, and those are easily calculated for proteins based on their published PDB structures [7,51]. This anomalous diffusion is thus a direct signature of the fractal structure of proteins. The result for large fractals is a stretched exponential decay of the dynamic structure factor, $S(k,t) \approx S(k)\exp[-(\Gamma_k t)^{\beta}]$, with anomalous k dependence of the relaxation rate, $\Gamma_k \sim k^{2/\beta}$, and a stretching exponent β that identifies with the anomalous diffusion exponent ν. Besides proteins, our theory can be used for other systems exhibiting fractal structures, e.g., colloidal gels [52], chromatin [53], and colloidal glasses [54].

The anomalous wave-number dependence of the relaxation rate, $\Gamma_k \sim k^{2/\nu}$, and the anomalous diffusion, $\langle \Delta\vec{u}(t)^2\rangle \sim t^{\nu}$, can be explained using simple scaling hypotheses. For the relaxation rate we assume $\Gamma_k = Dk^2 h(k\bar{u})$, where $h(x)$ is a scaling function and D is the center-of-mass diffusion coefficient of the fractal. Note that the scaling variable is $k\bar{u}$ rather than kR_g, as for flexible polymers. We make use of the generalized Landau-Peierls instability, $\bar{u} \sim N^{1/d_s - 1/2}$, and take $D \sim N^{-1}$ and $D \sim R_g^{-1} \sim N^{-1/d_f}$ for the Rouse- and Zimm-type models of friction [33], respectively. Demanding that Γ_k is independent of N for $k\bar{u} \gg 1$, the scaling function for $x \gg 1$ must satisfy $h(x) \sim x^{2d_s/(2-d_s)}$ (for the Rouse model) and $h(x) \sim x^{2d_s/[d_f(2-d_s)]}$ (for the Zimm model), leading to $\Gamma_k \sim k^{2/\nu}$ with $\nu = 1 - d_s/2$ (Rouse) and $\nu = (2-d_s)/(2-d_s+d_s/d_f)$ (Zimm) as stated. Similarly, for the MSD we assume $\langle \Delta\vec{u}(t)^2\rangle = \bar{u}^2\phi(t/\tau_N)$ where $\phi(x)$ is the scaling function and τ_N is the longest vibrational relaxation time. Assuming $\tau_N \simeq \bar{u}^2/D$, such that $\tau_N \sim N^{2/d_s}$ (Rouse) and $\tau_N \sim N^{2/d_s - 1 + 1/d_f}$ (Zimm), and demanding that for $t \ll \tau_N$ the MSD is independent of N, it follows that $\phi(x) \sim x^{\nu}$ for $x \ll 1$ (with ν taking the above stated values associated with the Rouse and Zimm models) leading to $\langle \Delta\vec{u}(t)^2\rangle \sim t^{\nu}$.

Numerical evaluation for proteins in solutions demonstrates that the small size of most proteins, combined with the influence of rotational and translational diffusion processes, leads to significant deviations of the value of β from its "infinite network" limit ν. Nevertheless, β remains a fingerprint of the fractal nature of proteins. Moreover, ν influences the value of β, and serves as an effective lower bound for it. Comparison with recent neutron spin-echo studies on hemoglobin and

myoglobin is not entirely satisfactory and may motivate further studies in this direction. As a practical conclusion we suggest performing such experiments with proteins bounded to a surface [55,56] similarly to what was done in Ref. [57]. This will diminish to a minimum the contribution from translational and rotational degrees of freedom and will thus allow for better exploration of internal degrees of freedom.

ACKNOWLEDGMENTS

R.G. and J.K. acknowledge support from the Deutsch-Israelische Projektkooperation (DIP). R.G. acknowledges partial support from the National Science Foundation under Grant No. PHY05-51164 and thanks the Kavli Institute for Theoretical Physics at UCSB for hospitality under the program "Biological Frontiers of Polymer and Soft Matter Physics," where part of this work was done. S.R. acknowledges support from the Converging Technologies program of the Israeli Council for higher education.

APPENDIX A: PROOF OF EQ. (39)

1. Disorder-averaged correlation functions

To prove Eq. (39) we develop the correlation functions in this equation. Note that here it is implicitly assumed that disorder average (a spatial average in which we sum over $\vec{\ell}$ and $\vec{\ell}'$ and divide by N^2) is being performed in addition to the thermal (time) average. This eliminates any real positional dependence, leaving only the relative distance between the two amino acids as a variable. It is easy to show that

$$\begin{aligned}&\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle\\&\quad=\langle\vec{u}(\vec{\ell},t)^2\rangle+\langle\vec{u}(\vec{\ell}',0)^2\rangle-2\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell}',0)\rangle\\&\quad=2\langle\vec{u}^{\,2}\rangle-2\langle\vec{u}(|\vec{\ell}-\vec{\ell}'|,t)\cdot\vec{u}(0,0)\rangle \qquad \text{(A1)}\\&\quad\langle\Delta\vec{u}(t)^2\rangle=2\langle\vec{u}^{\,2}\rangle-2\langle\vec{u}(0,t)\cdot\vec{u}(0,0)\rangle \qquad \text{(A2)}\\&\langle\vec{x}(t)\cdot\vec{x}(0)\rangle\\&\quad=\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',t)]\cdot[\vec{u}(\vec{\ell},0)-\vec{u}(\vec{\ell}',0)]\rangle\\&\quad=\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell},0)\rangle+\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell}',0)\rangle\\&\qquad-\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell}',0)\rangle-\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell},0)\rangle\\&\quad=2\langle\vec{u}(0,t)\cdot\vec{u}(0,0)\rangle-2\langle\vec{u}(|\vec{\ell}-\vec{\ell}'|,t)\cdot\vec{u}(0,0)\rangle. \qquad \text{(A3)}\end{aligned}$$

Substituting the above three equations into Eq. (39) leads to an identity.

2. Exact correlation functions

Here we derive an expression that is equivalent to Eq. (39) using exact correlation functions. To do so, first consider the combination of two-point correlation functions $\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle+\langle[\vec{u}(\vec{\ell}',t)-\vec{u}(\vec{\ell},0)]^2\rangle$ that is symmetric for exchange between $\vec{\ell}$ and $\vec{\ell}'$ (regardless of the present, model-dependent, symmetry of each one),

$$\begin{aligned}&\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle+\langle[\vec{u}(\vec{\ell}',t)-\vec{u}(\vec{\ell},0)]^2\rangle\\&\quad=\langle\vec{u}(\vec{\ell},t)^2\rangle+\langle\vec{u}(\vec{\ell}',t)^2\rangle+\langle\vec{u}(\vec{\ell},0)^2\rangle+\langle\vec{u}(\vec{\ell}',0)^2\rangle\\&\qquad-2\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell}',0)\rangle-2\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell},0)\rangle\\&\quad=2\langle\vec{u}(\vec{\ell})^2\rangle+2\langle\vec{u}(\vec{\ell}')^2\rangle-2\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell}',0)\rangle\\&\qquad-2\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell},0)\rangle. \qquad \text{(A4)}\end{aligned}$$

Next, consider the MSDs of the specific (non-disorder-averaged) amino acids $\vec{\ell}$ and $\vec{\ell}'$,

$$\langle\Delta\vec{u}(\vec{\ell},t)^2\rangle=2\langle\vec{u}(\vec{\ell})^2\rangle-2\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell},0)\rangle, \qquad \text{(A5)}$$

$$\langle\Delta\vec{u}(\vec{\ell}',t)^2\rangle=2\langle\vec{u}(\vec{\ell}')^2\rangle-2\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell}',0)\rangle. \qquad \text{(A6)}$$

Finally, consider the autocorrelation function of the fluctuation, $\vec{x}(\vec{\ell},\vec{\ell}',t)$, in the vector of separation between the specific amino acids $\vec{\ell}$ and $\vec{\ell}'$,

$$\begin{aligned}&\langle\vec{x}(\vec{\ell},\vec{\ell}',t)\cdot\vec{x}(\vec{\ell},\vec{\ell}',0)\rangle\\&\quad=\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',t)]\cdot[\vec{u}(\vec{\ell},0)-\vec{u}(\vec{\ell}',0)]\rangle\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell},0)\rangle\\&\qquad+\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell}',0)\rangle-\langle\vec{u}(\vec{\ell},t)\cdot\vec{u}(\vec{\ell}',0)\rangle-\langle\vec{u}(\vec{\ell}',t)\cdot\vec{u}(\vec{\ell},0)\rangle. \qquad \text{(A7)}\end{aligned}$$

Combining Eqs. (A4)–(A7), and using now the symmetry for exchange between $\vec{\ell}$ and $\vec{\ell}'$, $\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle=\langle[\vec{u}(\vec{\ell}',t)-\vec{u}(\vec{\ell},0)]^2\rangle$, that follows from Eq. (21), leads to the following identity:

$$\begin{aligned}&\langle[\vec{u}(\vec{\ell},t)-\vec{u}(\vec{\ell}',0)]^2\rangle\\&\quad=\tfrac{1}{2}[\langle\Delta\vec{u}(\vec{\ell},t)^2\rangle+\langle\Delta\vec{u}(\vec{\ell}',t)^2\rangle]+\langle\vec{x}(\vec{\ell},\vec{\ell}',t)\cdot\vec{x}(\vec{\ell},\vec{\ell}',0)\rangle. \qquad \text{(A8)}\end{aligned}$$

APPENDIX B: EFFECT OF ROTATIONAL DIFFUSION

Here we evaluate approximately the effect of protein rotations on the dynamic structure factor for short times. An alternative, more complete, calculation, can be found in Ref. [46]; however, the approximate calculation below suffices for our purposes. Rotations result in a change of the angle between the wave vector $\vec{k}$ and the vector of each residue i, $\vec{R}_i(t)$, in the protein center-of-mass coordinate frame. At $t=0$ this angle is a random variable, changing from one protein to the other within the scattering ensemble. As a result of protein rotational diffusion, this angle is changing with time, and we can describe this change by the unit vector $\hat{n}_i(t)$ that denotes the direction of the vector $\vec{R}_i(t)$, $\vec{R}_i(t)=R_i\hat{n}_i(t)$, where it is understood that the dependence on time in $\vec{R}_i(t)$ is *solely due to rotations*, as vibrations are included separately. Consider the average

$$\langle\cdots\rangle\equiv\left\langle e^{i\vec{k}\cdot[\vec{R}_i(t)-\vec{R}_j(0)]}\right\rangle. \qquad \text{(B1)}$$

Accordingly we write $\vec{R}_i(t)=\vec{R}_i(0)+R_i[\hat{n}_i(t)-\hat{n}_i(0)]$ leading to

$$\langle\cdots\rangle=e^{i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)}\left\langle e^{iR_i\vec{k}\cdot[\hat{n}_i(t)-\hat{n}_i(0)]}\right\rangle, \qquad \text{(B2)}$$

where, for brevity, we have defined $\vec{R}_i=\vec{R}_i(0)$ and $\vec{R}_j=\vec{R}_j(0)$.

At early times, $D_{\rm rot}t\ll1$, where $D_{\rm rot}$ is the rotational diffusion coefficient, the vector $\vec{s}_i(t)=\hat{n}_i(t)-\hat{n}_i(0)$ is almost perpendicular to $\hat{n}_i(0)$, being almost tangent to the unit sphere at $\hat{n}_i(0)$. Its radial component, along $\hat{n}_i(0)$ is negligible. Hence this two-dimensional vector represent a two-dimensional (almost) regular diffusion, and its two components are independent and obey almost exactly Gaussian statistics. We denote by $\vec{k}_t$ the two-dimensional projection of $\vec{k}$ on the plane that is tangent to the unit sphere at $\hat{n}_i(0)$, so that the magnitude

of $\vec{k}_t$ is $k_t = k\sin\theta_i$ where θ_i is the angle between $\vec{k}$ and R_i. Using these definitions

$$\langle\cdots\rangle \simeq e^{i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)}\langle e^{iR_i\vec{k}_t\cdot\vec{s}_i(t)}\rangle. \tag{B3}$$

Thus, using the nearly Gaussian statistics property of the two-dimensional vector $\vec{s}_i(t)$,

$$\langle\cdots\rangle \simeq e^{i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)}e^{-\frac{1}{4}k_i^2R_i^2\langle\vec{s}_i(t)^2\rangle}. \tag{B4}$$

Here we require the calculation of the unit vector MSD

$$\langle\vec{s}_i(t)^2\rangle = \langle[\hat{n}_i(t)-\hat{n}_i(0)]^2\rangle = 2[1-\langle\hat{n}_i(t)\cdot\hat{n}_i(0)\rangle]. \tag{B5}$$

In rotational diffusion theory one finds [33]

$$\langle\hat{n}_i(t)\cdot\hat{n}_i(0)\rangle = e^{-2D_{\rm rot}t}, \tag{B6}$$

where $D_{\rm rot}$ is the rotational diffusion coefficient. Hence

$$\langle[\hat{n}_i(t)-\hat{n}_i(0)]^2\rangle = 2(1-e^{-2D_{\rm rot}t}) \simeq 4D_{\rm rot}t, \tag{B7}$$

where the last equality holds for short times $D_{\rm rot}t \ll 1$. This shows that rotation angle $\delta\Omega(t)$ between $\hat{n}_i(t)$ and $\hat{n}_i(0)$ obeys $\langle\delta\Omega(t)^2\rangle \simeq 4D_{\rm rot}t$.

Using these results in the dynamic structure factor that now includes both rotations [as $\vec{R}_i(t)$] and vibrations

$$S_p(\vec{k},t) \simeq \frac{1}{N}\sum_{i,j}\langle e^{i\vec{k}\cdot[\vec{R}_i(t)-\vec{R}_j(0)]}\rangle e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle} \tag{B8}$$

leads to

$$S_p(\vec{k},t) \simeq \frac{1}{N}\sum_{i,j} I(\vec{k},t)e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}, \tag{B9}$$

where

$$I(\vec{k},t) = e^{ikR_{ij}\cos\theta_{ij}}\exp\big[-k^2R_i^2\sin^2\theta_i D_{\rm rot}t\big], \tag{B10}$$

and θ_{ij} is the angle between $\vec{k}$ and $\vec{R}_{ij} \equiv \vec{R}_i - \vec{R}_j$. Performing the angular average over the initial angles is equivalent to rotating the wave vector $\vec{k}$ in all directions, yielding $\langle I(\vec{k},t)\rangle_{\vec{k}}$ and thus $\langle S_p(\vec{k},t)\rangle_{\vec{k}}$. Since one has to keep the angle between the vectors R_i and R_j fixed, this implies that the angles θ_i and θ_{ij} are dependent. Therefore, for accurate analysis we turn to numerical evaluation. The angularly averaged dynamics structure factor (excluding translational diffusion) is thus

$$S_p(k,t) = \frac{1}{N}\sum_{i,j}\langle I(\vec{k},t)\rangle_{\vec{k}}e^{-\frac{k^2}{6}\langle[\vec{u}_i(t)-\vec{u}_j(0)]^2\rangle}. \tag{B11}$$

Equation (B11) allows to estimate the effect of rotational diffusion at short times, $D_{\rm rot}t \ll 1$.

APPENDIX C: PROTEIN STRUCTURE FACTOR: DERIVATION OF APPROXIMATE RESULTS

Here we evaluate approximately Eq. (68) using the two-point correlation function derived in Sec. III. As Eqs. (52)–(53) already involve positional averaging, we can ignore details of the protein structure.

1. Static structure factor

Consider first the static structure factor

$$S_p(k) = \frac{1}{N}\sum_{i,j=1}^{N}\frac{\sin[kR_{ij}]}{kR_{ij}}\exp\Big[-\frac{k^2}{6}C\frac{k_BT}{m\omega_o^2}\left(\frac{R_{ij}}{b}\right)^{d_w-d_f}\Big]. \tag{C1}$$

To examine the effect of vibrations on the static structure factor we also define a "frozen network" structure factor that does not include the vibrational contributions

$$\tilde{S}_p(k) = \frac{1}{N}\sum_{i,j}\frac{\sin[kR_{ij}]}{kR_{ij}}. \tag{C2}$$

If the network is assumed "infinitely" large we can ignore [as in Eqs. (52)–(53)] boundary effects and perform the summation over j. Transforming the sum to an integral we can formally write

$$S_p(k) = \frac{1}{b^{d_f}}\int_{V_g} d^{d_f}r\frac{\sin[kr]}{kr}\exp\Big[-\frac{k^2}{6}C\frac{k_BT}{m\omega_o^2}\left(\frac{r}{b}\right)^{d_w-d_f}\Big] \tag{C3}$$

and similarly for $\tilde{S}_p(k)$ (with the second, exponential, term in the integrand equal to unity). For $kR_g \gg 1$ and in case of a "frozen network" this leads to the well-known result $\tilde{S}_p(k) \sim k^{-d_f}$, as can be also obtained by scaling approach. More precisely,

$$\tilde{S}_p(k) \simeq \frac{\pi^{(d_f+1)/2}2^{d_f-1}}{\Gamma\big[\frac{3}{2}-\frac{d_f}{2}\big]b^{d_f}}\,k^{-d_f}. \tag{C4}$$

Yet some contribution, even if small, is obtained from the vibrations, similar to the broadening of the x-ray Brag peaks. Expanding the second term in the integrand of Eq. (C3) to leading order we obtain the first-order correction to $S_p(k) \sim k^{-d_f}$ in the form

$$S_p(k) \approx \tilde{S}_p(k) - \frac{Cd_f\pi^{(d_f+1)/2}2^{d_w-1}\Gamma[1+d_w/2]}{6d_w\Gamma[1+d_f/2]\Gamma\big[\frac{3}{2}-\frac{d_w}{2}\big]}\frac{k_BT}{m\omega_o^2b^{d_w}}\,k^{2-d_w} + \cdots. \tag{C5}$$

2. Dynamic structure factor: Vibrations

Next we consider the dynamic structure factor. It will be later shown that the rotational diffusion time dependence in $S_p(q,t)$ is rather weak for $t \lesssim \tau_N$. Taking this as an assumption for now, the dominant relaxation is due to vibrations. For the first level of analysis, we shall therefore ignore rotational diffusion and use Eq. (67).

For short times, $t \ll \tau(k)$, where $\tau(k) = t^*(r = k^{-1}) \simeq A^{-1}\omega_o^{-\theta}(kb)^{-d_f\theta/d_s}$, we can use Eq. (52) to find that the dynamic structure factor did not decay much and is left almost equal to the static structure factor, $S_p(k,t) \simeq S_p(k)$. A more accurate description can be obtained as follows. For all pairs for which "information" has already propagated between them, $R_{ij} < \xi(t) < 1/k$, we should use Eq. (53). For all pairs for which information has not yet arrived, $\xi(t) < R_{ij}$, we should

use Eq. (52). Thus

$$S_p(k,t) = \frac{1}{N}\sum_{i,j;R_{ij}<\xi(t)}^{N} \frac{\sin[kR_{ij}]}{kR_{ij}} \exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]\exp\left[-\frac{k^2}{6}C_1\frac{k_BT}{m\omega_o^{d_s}}\,(R_{ij}/b)^{2d_f/d_l}\,(\bar{A}t)^{-\mu}\right]$$

$$+\frac{1}{N}\sum_{i,j;R_{ij}>\xi(t)}^{N} \frac{\sin[kR_{ij}]}{kR_{ij}} \exp\left[-\frac{k^2}{6}C\frac{k_BT}{m\omega_o^2}\left(\frac{R_{ij}}{b}\right)^{d_w-d_f}\right]. \tag{C6}$$

It appears as a very good approximation to set the third, exponential, term in the sum running on $R_{ij} < \xi(t)$ to unity, since its argument is much smaller than one in the relevant timescales. Transforming the sum running on pairs with $R_{ij} < \xi(t)$ to an integral and evaluating the integral approximately for $k\xi(t) \ll 1$, Eq. (67) becomes

$$S_p(k,t) \simeq \frac{\pi^{d_f/2}}{\Gamma[d_f/2+1]b^{d_f}}\xi(t)^{d_f}\left[1-\frac{d_f}{6(2+d_f)}k^2\xi(t)^2+\cdots\right]\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]$$

$$+\frac{1}{N}\sum_{i,j;R_{ij}>\xi(t)}^{N} \frac{\sin[kR_{ij}]}{kR_{ij}} \exp\left[-\frac{k^2}{6}C\frac{k_BT}{m\omega_o^2}\left(\frac{R_{ij}}{b}\right)^{d_w-d_f}\right]. \tag{C7}$$

Equivalently,

$$S_p(k,t) \simeq S_p(k) - \frac{\pi^{d_f/2}}{\Gamma[d_f/2+1]b^{d_f}}\xi(t)^{d_f}\left(1-\frac{d_f}{6(2+d_f)}k^2\xi(t)^2+\cdots\right)\left(1-\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]\right), \tag{C8}$$

which, using $\xi(t) \simeq \ell(t)^{d_l/d_f} \simeq \omega_o^{d_s/d_f}(\bar{A}t)^{\frac{d_s}{d_f\theta}}$ and expanding for short times (using $\nu + d_s/\theta = 2/\theta$), yields, to leading order,

$$S_p(k,t) \simeq S_p(k) - C_3\,k^2\,t^{2/\theta}, \tag{C9}$$

where

$$C_3 = \frac{\pi^{d_f/2}}{6\Gamma[d_f/2+1]b^{d_f}}\omega_o^{d_s}\bar{A}^{\frac{d_s}{\theta}}B. \tag{C10}$$

Note that for the Rouse model ($\theta = 2$) this implies that the short time decay is roughly exponential.

At longer times, $\tau(k) \ll t \ll \tau_N$, i.e., when $1 \ll k\xi(t) \ll kR_g$ such that information has propagated beyond the scattering wavelength $\sim 1/k$, we can use Eq. (53) in Eq. (67). This is true even though the sum in Eq. (67) include all pairs; thus it includes also pairs for which information has not yet propagated between them and hence obeying Eq. (52). However, these pairs are distant much further than $1/k$ apart, $k^{-1} \ll \xi(t) \ll R_{ij}$; hence they contribute a vanishingly small value to the sum due to the familiar property of the Fourier transform. Physically, this implies that the "blob" of size $\sim 1/k$, which is controlling the relaxation at wave vector $\vec{k}$, is moving almost together as if it was a single bead. This leads [setting again the third, exponential, term in the sum running on $R_{ij} < \xi(t)$ to unity] to a *stretched exponential* decay of the dynamic structure factor,

$$S_p(k,t) \simeq \tilde{S}_p(k)\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right], \tag{C11}$$

where $\tilde{S}_p(k)$ is the "frozen fractal" static structure factor. As shown above, for not too soft vibrations $\tilde{S}_p(k) \simeq S_p(k)$. Equivalently, this result may be written as

$$S_p(k,t) \simeq \tilde{S}_p(k)\exp\left[-\frac{k^2}{6}\langle\Delta\vec{u}(t)^2\rangle\right], \tag{C12}$$

or, more conveniently, as

$$S_p(k,t) \simeq \tilde{S}_p(k)\exp[-(\Gamma_k t)^{\nu}], \tag{C13}$$

where

$$\Gamma_k = (B/6)^{1/\nu}k^{2/\nu}. \tag{C14}$$

The stretched exponential decay, together with the dependence of the stretching exponent ν, is thus a strong signature of the fractal structure.

3. Dynamic structure factor: Vibrations and rotational diffusion

To estimate the corrections to this result due to rotational diffusion [cf. Eq. (68)], we first introduce a decoupling approximation to the angular average of $\langle I(\vec{k},t)\rangle$ [Eq. (B10)],

$$\langle I(\vec{k},t)\rangle_{\vec{k}} \simeq \langle e^{ikR_{ij}\cos\theta_{ij}}\rangle_{\vec{k}}\left\langle\exp\left[-k^2R_i^2\sin^2\theta_i D_{\rm rot}t\right]\right\rangle_{\vec{k}}, \tag{C15}$$

in which we ignore the coupling between the angles θ_{ij} and θ_i. Thus we can use

$$\langle e^{ikR_{ij}\cos\theta_{ij}}\rangle_{\vec{k}} = \frac{\sin[kR_{ij}]}{kR_{ij}} \tag{C16}$$

and

$$\left\langle\exp\left[-k^2R_i^2\sin^2\theta_i D_{\rm rot}t\right]\right\rangle_{\vec{k}} = \frac{F[kR_i\sqrt{D_{\rm rot}t}]}{kR_i\sqrt{D_{\rm rot}t}}, \tag{C17}$$

where $F[x]$ is the Dawson integral defined by $F(x) = e^{-x^2}\int_0^x dy\, e^{y^2}$. Next we take the "infinite" fractal limit, and

replace sums by integrals, to obtain

$$S_p(k,t) \simeq \frac{1}{N}\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]\frac{1}{b^{2d_f}}\int_{V_g} d^{d_f}r\int_{V_g} d^{d_f}r'\frac{\sin[k|\vec{r}-\vec{r}\,'|]}{k|\vec{r}-\vec{r}\,'|}\frac{F[kr'\sqrt{D_{\rm rot}t}]}{kr'\sqrt{D_{\rm rot}t}}$$
$$\times\exp\left[-\frac{k^2}{6}C_1\frac{k_BT}{m\omega_o^{d_s}}\,(|\vec{r}-\vec{r}\,'|/b)^{2d_f/d_l}\,(\bar{A}t)^{-\mu}\right]. \tag{C18}$$

We now assume *very large* k such that $kR_g\sqrt{D_{\rm rot}t}\gg 1$ (although our short time assumption $D_{\rm rot}t\ll 1$ still holds). Integrating over r' leads to

$$S_p(k,t) \simeq \frac{d_f\pi^{3/2+d_f/2}\mathrm{Csc}[d_f\pi/2]}{4Nb^{2d_f}\Gamma[3/2-d_f/2]\Gamma[d_f/2+1]D_{\rm rot}^{d_f/2}}k^{-d_f}\,t^{-d_f/2}\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]$$
$$\times\int_{V_g} d^{d_f}r\frac{\sin[kr]}{kr}\exp\left[-\frac{k^2}{6}C_1\frac{k_BT}{m\omega_o^{d_s}}\,(r/b)^{2d_f/d_l}\,(\bar{A}t)^{-\mu}\right]. \tag{C19}$$

Integrating over r, and expanding the second, exponential, term in the integrand to leading order, we find

$$S_p(k,t)\approx \text{const}\times k^{-d_f}\,t^{-d_f/2}\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]\left[\tilde{S}_p(k)-\text{const}\times\frac{k_BT}{m\omega_o^{d_s}b^{2d_f/d_l}}k^{2-d_f-2d_f/d_l}(\bar{A}t)^{-\mu}\right]. \tag{C20}$$

Thus we find that, to leading order, the dynamics structure factor decays according to

$$S_p(k,t)\approx t^{-d_f/2}\exp\left(-\frac{1}{6}k^2Bt^{\nu}\right)\approx t^{-d_f/2}\exp[-(\Gamma_k t)^{\nu}]. \tag{C21}$$

In the case of not so large k or very short times, such that $kR_g\sqrt{D_{\rm rot}t}\ll 1$ (but still $kR_g\gg 1$), we find, to leading order,

$$S_p(k,t)\approx\exp\left[-\frac{k^2}{6}B\,t^{\nu}\right]\left[1-\frac{2d_f}{3(2+d_f)}k^2R_g^2D_{\rm rot}t\right]\left[\tilde{S}_p(k)-\text{const}\times\frac{k_BT}{m\omega_o^{d_s}b^{2d_f/d_l}}k^{2-d_f-2d_f/d_l}(\bar{A}t)^{-\mu}\right]. \tag{C22}$$

Assuming $R_g\gg b$ we have $Bt^{\nu}\gg(k_BT/\eta R_g)t$ in the relevant time regime $t\ll\tau_N$. In this case we find, to leading order, a pure stretched exponential decay,

$$S_p(k,t)\approx\exp\left(-\frac{1}{6}k^2Bt^{\nu}\right)\approx\exp[-(\Gamma_k t)^{\nu}]. \tag{C23}$$

[1] R. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).
[2] M. B. Enright and D. M. Leitner, Phys. Rev. E **71**, 011912 (2005).
[3] X. Yu and D. M. Leitner, J. Chem. Phys. **119**, 12673 (2003); D. M. Leitner, Annu. Rev. Phys. Chem. **59**, 233 (2008).
[4] R. Burioni, D. Cassi, F. Cecconi, and A. Vulpiani, Proteins: Struct. Func. Bioinf. **55**, 529 (2004); R. Burioni, D. Cassi, M. P. Fontana, and A. Vulpiani, Europhys. Lett. **58**, 806 (2002).
[5] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[6] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008).
[7] M. de Leeuw, S. Reuveni, J. Klafter, and R. Granek, PLoS ONE **4**, e7296 (2009).
[8] S. Reuveni, R. Granek, and J. Klafter, Proc. Natl. Acad. Sci. **107**, 13696 (2010).
[9] A. Banerji and I. Ghosh, PLoS ONE **4**, e7361 (2009).
[10] R. Granek, Phys. Rev. E **83**, 020902(R) (2011).
[11] H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinson, and S. R. Kurtz, Phys. Rev. Lett. **45**, 1456 (1980).
[12] S. G. Lushnikov, A. V. Svanidze, and I. L. Sashin, JETP Lett. **82**, 30 (2005).
[13] For a review, see A. Banerji and I. Ghosh, Cell. Mol. Life Sci. **68**, 2711 (2011).
[14] S. C. Kou and X. S. Xie, Phys. Rev. Lett. **93**, 180603 (2004); W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, *ibid.* **94**, 198302 (2005).
[15] P. Senet, G. G. Maisuradze, C. Foulie, P. Delarue, and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **105**, 19708 (2008).
[16] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, Phys. Rev. Lett. **100**, 188103 (2008).
[17] M. T. Zimmermann, A. Skliros, A. Kloczkowski, and R. L. Jernigan, Immunome Res. **7**, 5 (2011).
[18] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1992).
[19] For a review, see T. Nakayama, K. Yakubo, and R. L. Orbach, Rev. Mod. Phys. **66**, 381 (1994).
[20] S. Alexander and R. Orbach, J. Phys. Lett. **43**, L625 (1982).
[21] S. Alexander, Phys. Rev. B **40**, 7953 (1989).
[22] H. Morita and M. Takano, Phys. Rev. E **79**, 020901(R) (2009).
[23] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997); I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Phys. Rev. Lett. **80**, 2733 (1998).
[24] M. E. Cates, J. Phys. (France) **46**, 1059 (1985).
[25] M. Monkenbusch and D. Richter, C. R. Phys. **8**, 845 (2007).
[26] M. Monkenbusch, D. Richter, and R. Biehl, Chem. Phys. Chem. **8**, 1188 (2010).

[27] J. Lal, P. Fouquet, M. Maccarini, and L. Makowski, J. Mol. Biol. **397**, 423 (2010).
[28] J. H. Roh, J. E. Curtis, S. Azzam, V. N. Novikov, I. Peral, Z. Chowdhuri, R. B. Gregory, and A. P. Sokolov, Biophys. J. **91**, 2573 (2006).
[29] S. Khodadadi, S. Pawlus, J. H. Roh, V. Garcia Sakai, E. Mamontov, and A. P. Sokolov, J. Chem. Phys. **128**, 195106 (2008).
[30] L. Hong, N. Smolin, B. Lindner, A. P. Sokolov, and J. C. Smith, Phys. Rev. Lett. **107**, 148102 (2011).
[31] G. R. Kneller and K. Hinsen, J. Chem. Phys. **121**, 10278 (2004).
[32] V. Calandrini, V. Hamon, K. Hinsen, P. Calligari, M.-C. Bellissent-Funel, and G. R. Kneller, Chem. Phys. **345**, 289 (2008).
[33] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
[34] A. G. Zilman and R. Granek, Phys. Rev. E **58**, R2725 (1998).
[35] A. Blumen, C. von Ferber, A. Jurjiu, and T. Koslowski, Macromolecules **37**, 638 (2004); A. Blumen, A. Jurjiu, T. Koslowski, and C. von Ferber, Phys. Rev. E **67**, 061103 (2003).
[36] A. Bunde, H. E. Roman, S. Russ, A. Aharony, and A. B. Harris, Phys. Rev. Lett. **69**, 3189 (1992).
[37] J. W. Kantelhardt and A. Bunde, Phys. Rev. E **56**, 6693 (1997).
[38] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. E **81**, 040103(R) (2010).
[39] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. E **82**, 041132 (2010).
[40] L.-W. Yang, E. Eyal, C. Chennubhotla, J. JunGoo, A. M. Gronenborn, and I. Bahar, Structure **15**, 741 (2007).
[41] B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley, Malabar, 1990).
[42] T. Takedaa, Y. Kawabata, H. Seto, S. Komura, S. K. Ghosh, M. Nagao, and D. Okuhara, J. Phys. Chem. Solids **60**, 1375 (1999); S. Komura, T. Takeda, H. Seto, and M. Nagao, in *Neutron Spin Echo Spectroscopy: Basics, Trends and Applications*, edited by F. Mezei, C. Pappas, and T. Gutberlet (Springer, Berlin, 2003), pp. 302–311.
[43] P. Falus, M. A. Borthwick, and S. G. J. Mochrie, Phys. Rev. Lett. **94**, 016105 (2005); P. Falus, M. A. Borthwick, S. Narayanan, A. R. Sandy, and S. G. J. Mochrie, *ibid.* **97**, 066102 (2006).
[44] The hydrodynamic radius may be estimated using the Kirkwood-Riseman formula [33,45] $R_h^{-1} = \langle R_{ij}^{-1} \rangle$, where $R_{ij} = |\vec{R}_i - \vec{R}_j|$.
[45] W. Van Saarloos, Physica A **147**, 280 (1987); M. Lattuada, H. Wu, and M. Morbidelli, J. Colloid Interf. Sci. **268**, 96 (2003).
[46] H. M. Lindsay, R. Klein, D. A. Weitz, M. Y. Lin, and P. Meakin, Phys. Rev. A **38**, 2614 (1988).
[47] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[48] G. Stokes, Trans. Cambridge Philos. Soc. **9**, 5 (1856); A. Einstein, Ann. Phys. (Leipzig) **19**, 371 (1906); P. Debye, *Polar Molecules* (Dover, New York, 1929).
[49] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).
[50] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1988).
[51] For a data set of d_s and d_f values containing about 5000 proteins see the supporting information in Ref. [7].
[52] A. H. Krall and D. A. Weitz, Phys. Rev. Lett. **80**, 778 (1998).
[53] E. L. Aiden *et al.*, Science **326**, 289 (2009); A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. (France) **49**, 2095 (1988); A. Y. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. **23**, 373 (1993); J. G. McNally and D. Mazza, EMBO J. **29**, 2 (2010); A. J. Einstein, H.-S. Wu, M. Sanchez, and J. Gil, J. Pathol. **185**, 366 (1998); V. Bedin, R. L. Adam, B. C. S. de Sà, G. Landman, and K. Metze, BMC Cancer **10**, 260 (2010).
[54] J. Mattsson, H. M. Wyss, A. Fernandez-Nieves, K. Miyazaki, Z. Hu, D. R. Reichman, and D. A. Weitz, Nature (London) **462**, 83 (2009).
[55] A. M. Klibanov, Science **219**, 722 (1983).
[56] E. K. Katzir, Trends Biotechnol. **11**, 471 (1993).
[57] M. Tehei, J. C. Smith, C. Monk, J. Ollivier, M. Oettl, V. Kurkal, J. L. Finney, and R. M. Daniel, Biophys. J. **90**, 1090 (2006).

PHYSICAL REVIEW E **81**, 040103(R) (2010)

Vibrational shortcut to the mean-first-passage-time problem

Shlomi Reuveni,[1,2] Rony Granek,[3] and Joseph Klafter[1,*]
[1]*School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel*
[2]*School of Mathematics, Tel-Aviv University, Tel-Aviv 69978, Israel*
[3]*The Stella and Avram Goren-Goldstein Department of Biotechnology Engineering, Ben-Gurion University, Beer Sheva 84105, Israel*
(Received 8 November 2009; published 26 April 2010)

What is the average time a random walker takes to get from A to B on a fractal structure and how does this mean time scale with the size of the system and the distance between source and target? We take a nonprobabilistic approach toward this problem and show how the solution is readily obtained using an analysis of thermal vibrations on fractals. Invariance under scaling and continuity with respect to the spectral dimension are shown to be emergent properties of the solution obtained via vibrational analysis. Our result emphasizes the duality between diffusion and vibrations on fractal structures. Applications to biological systems are discussed.

DOI: 10.1103/PhysRevE.81.040103 PACS number(s): 05.40.Fb, 05.45.Df

In this Rapid Communication we consider a random walker moving on a fractal bounded domain of size N. We take a nonprobabilistic approach toward a probabilistic problem and find, based on the vibrational properties of the network, an approximate expression for the mean-first-passage time (MFPT) $T(N,r)$ between two distinct tagged points distanced r apart,

$$\frac{T(N,r)}{N} \simeq \tilde{C}\begin{cases} \dfrac{1-(r/a)^{d_f(2/d_s-1)}}{[d_s/2-1]} & d_s>2 \\ d_f \ln(r/a) & d_s=2 \\ \dfrac{-1+(r/a)^{d_f(2/d_s-1)}}{[1-d_s/2]} & d_s<2 \end{cases}. \qquad (1)$$

Here a is the distance between nearest neighbors, d_s and d_f are the network spectral and fractal dimensions, respectively, and $\tilde{C}$ is a constant to be defined later in the Rapid Communication. The joint domain in which the approximation is valid is given by: $\{N \gg 1, a \ll r \ll R_g\}$ where R_g is the radius of gyration. The spectral dimension d_s governs the density of low-frequency normal modes of a fractal. More precisely, denoting the density of modes $g(\omega)$, the scaling relation $g(\omega) \sim \omega^{d_s-1}$ holds for low frequencies. Describing the mass fractal dimension d_f is most convenient using a three-dimensional example. Draw a sphere of radius r enclosing some lattice points in space and calculate their mass $M(r)$, increase r, and calculate again. Do this several times and if $M(r)$ scales as r^{d_f} the exponent d_f is called the fractal dimension. We note that the ratio $d_s/2d_f$ is just the Hurst exponent of the random-walk mean-square displacement in time [1].

It is important to notice that in general the MFPT between two points on a finite, anisotropic, and inhomogeneous structure may depend specifically on the identity of the source and target points. Equation (1) was hence written for the disorder average (average over all pairs distanced r apart) of the MFPT. The main conclusion from Eq. (1) is that as a function of the source-target distance r the normalized MFPT either grows like a power law ($d_s<2$) or saturates toward a limiting constant ($d_s>2$). In its domain of validity Eq. (1) covers both the numerically dominant part and the nonanalytic part of the MFPT. For $d_s<2$ the nonanalytic part is the numerically dominant one, while for $d_s>2$ it is not. We note that a result similar to Eq. (1) was obtained via probabilistic methods in a recent paper by Condamin *et al.* [2]. Other relations between first-passage-time statistics and the spectral dimension can be found in [3–5].

Deriving Eq. (1), we start by solving the analogous vibrational problem using scaling arguments. We then continue with a more direct treatment which leads to additional insights regarding the solution. Consider an elastic network (EN) of masses coupled by harmonic springs in the framework of what is known as the scalar elasticity model or the Gaussian network model (GNM) when applied for proteins [6,7]. The same network can be thought of as a network of nodes connected by edges (which we will refer to as RN) and we will alternate between these representations as we go (see Fig. 1). Let us now couple the EN to a thermal bath and consider a MFPT problem on the RN where at every time step the random walker on the RN performs a random jump to a nearest neighbor without preference. Consider two nodes on the RN, i and j, characterized by an Euclidean distance r between them. What is the MFPT $T(i,j)$ for motion between sites i and j? [If the MFPT depends on the direction of travel, we define $T(i,j)$ to be the average for the two directions.] On the EN the distance between mass i and j is not fixed due to thermal fluctuations. From the way we have constructed the EN (Fig. 1) it follows that the equilibrium (ensemble-average) distance is r, and we will denote the instantaneous distance by r_{ij}. It turns out that there is a relation between the thermal variance in r_{ij} and the MFPT $T(i,j)$ [8],

$$\frac{T(i,j)}{N} = \frac{z\gamma}{6k_BT}\langle (r_{ij}-r)^2\rangle = \frac{z\gamma}{6k_BT}\langle \Delta r_{ij}^2\rangle, \qquad (2)$$

where γ is the network spring constant and z is the mean coordination number of a node in the RN. See Fig. 1 for an explanation and visual illustration of this result. One way to evaluate $T(i,j)$ is via direct evaluation of the thermal variance $\langle \Delta r_{ij}^2\rangle$. We return to this approach later on. Our current

*Corresponding author; klafter@post.tau.ac.il

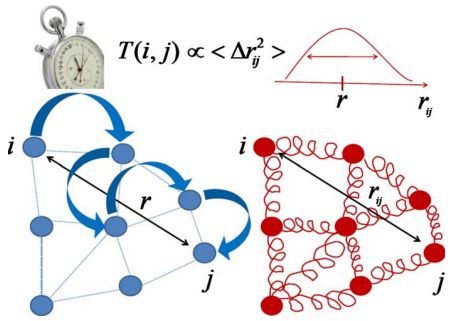


FIG. 1. (Color online) Left: a network of nodes connected by edges (RN); right: an analogs elastic network of masses coupled by harmonic springs (EN). One can transform a RN to an EN by transforming nodes to masses and edges to springs. We study the relation between a random walk on the RN and the thermal vibrations of the EN. Consider two nodes on the RN, i and j, characterized by an Euclidean distance r between them. On the EN the distance between mass i and j is not fixed due to thermal fluctuations. From the way we have constructed the EN it follows that the equilibrium (ensemble-average) distance is r, and we will denote the instantaneous distance by r_{ij}. The MFPT a random walker takes to get from node i to node j [if the MFPT depends on the direction of travel, we define $T(i,j)$ to be the average for the two directions] on a RN is proportional to the thermal variance in the distance between mass i and mass j on an EN coupled to a thermal bath.

efforts concentrate on an alternative route. A rearranged version of Eq. (2),

$$k_B T = \frac{Nz\gamma}{6T(i,j)}\langle \Delta r_{ij}^2 \rangle \equiv \frac{1}{2} k_{ij}^{eff} \langle \Delta r_{ij}^2 \rangle, \tag{3}$$

immediately lends itself to an effective inverse spring constant interpretation of $T(i,j)$. The higher the MFPT $T(i,j)$, the softer the effective spring connecting sites i and j; the softer the spring, the higher the magnitude of the fluctuations. We note here that a similar analogy was made between $T(i,j)$ and the point to point resistivity $\rho(i,j)$ in the analogous electrostatic problem [9,10]. Defining $T(N,r)$ as the average of $T(i,j)$ over all pairs with an equilibrium distance r, we conclude that the effective spring constant associated with this distance is

$$k_r^{eff} = \frac{Nz\gamma}{3T(N,r)}, \tag{4}$$

In deriving the scaling of k_r^{eff} for fractal structures we follow Alexander [11] and associate a length scale with a vibrational frequency. Consider the isolation of a fractal blob of size r from an original larger fractal. High-frequency modes for which $\omega \gg \omega(r)$ will not be effected by this isolation; on the other hand, low-frequency modes for which $\omega \ll \omega(r)$ will disappear from the spectrum due to the disconnection of the blob. By construction the crossover will occur at $\omega(r)$ the frequency we have associated with r. In the domain: $a \ll r \ll R_g$, this frequency was shown to scale as $\omega(r) \sim r^{-d_f/d_s}$ [11] and we note that this is nothing but the appropriate dispersion relation for a fractal elastic network. Let us now use this result in order to associate a length scale with an effective spring constant. Coarse graining the original fractal as we collapse all blobs of size r to points of mass $M(r) \sim r^{d_f}$, we construct a new network on this scale. Preserving the self-similarity we require that the basic frequency (highest) in the coarse-grained fractal would be $\omega(r)$, and it follows that the spring constant of the springs connecting the newly adjacent points should be [11]

$$k_r^{eff} \equiv M(r)\omega^2(r) \sim r^{d_f(1-2/d_s)}. \tag{5}$$

Substituting into Eq. (4) we obtain the correct scaling dependence of the normalized MFPT on the intersite distance,

$$\begin{cases} \dfrac{T(N,r)}{N} \sim r^{d_f(2/d_s-1)} & d_s \neq 2 \\ \dfrac{T(N,r)}{N} \sim \ln r & d_s = 2 \end{cases}, \tag{6}$$

where in the case of a vanishing exponent we have assumed a logarithmic correction. Equation (6) gives the dependence up to an unknown additive constant A and an unknown multiplicative constant B. The sign of the multiplicative constant B is determined by the rather intuitive requirement from the averaged MFPT to be a monotonically increasing function of the distance. It follows that the sign is a step function of the spectral dimension and that

$$\begin{cases} B < 0 & d_s > 2 \\ B > 0 & d_s \leq 2 \end{cases}. \tag{7}$$

We now provide a more direct derivation of the same result. This time we will start from Eq. (2) and directly evaluate $\langle \Delta r_{ij}^2 \rangle$. First we denote by $\vec{r}_i^{\,0}$ and $\vec{u}_i$ the equilibrium position of the ith mass and the instantaneous deviation from this position, respectively. It follows that the instantaneous distance separating any given pair is: $r_{ij} \equiv \sqrt{(\vec{r}_j^{\,0} + \vec{u}_j - \vec{r}_i^{\,0} - \vec{u}_i)^2}$. If the equilibrium distance between mass i and j is r it is straight forward to show that

$$\frac{T(i,j)}{N} = \frac{z\gamma}{6k_B T}\langle (r_{ij} - r)^2 \rangle = \frac{z\gamma}{6k_B T}\langle \vec{u}_j^2 + \vec{u}_i^2 - 2\vec{u}_i \cdot \vec{u}_j \rangle, \tag{8}$$

using the fact that $\langle r_{ij} \rangle = r$, and $\langle \vec{u}_j \rangle = \langle \vec{u}_i \rangle = 0$. In order to proceed we average over all the N_r pairs distanced r apart and get

$$\frac{T(r)}{N} \simeq \frac{z\gamma}{3k_B T}\left[\frac{1}{N}\sum_{i=1}^{N} \langle \vec{u}_i^2 \rangle - \frac{1}{N_r} \sum_{pairs} \langle \vec{u}_i \cdot \vec{u}_j \rangle \right] \tag{9}$$

where we have further assumed that the pairs average over the first two terms in the right-hand side of Eq. (8) is well represented by the average mean-square displacement [first term in Eq. (9)]. Calculating the thermal average we may consider the contribution of each normal mode separately. We consider two limits [12,13], normal modes with frequencies $\omega \ll \omega(r)$ result in a correlated movement (on average) of pairs separated r apart ($\vec{u}_i \cdot \vec{u}_j \simeq \vec{u}_i^2$), and hence in this limit,

$$\frac{1}{N}\sum_{i=1}^{N}\langle \vec{u}_i^2\rangle - \frac{1}{N_{r\,pairs}}\sum \langle \vec{u}_i \cdot \vec{u}_j\rangle \simeq 0. \quad (10)$$

Conversely, for normal modes with frequencies $\omega \gg \omega(r)$, the terms in the second sum add up incoherently and this term can hence be neglected. In this limit we get

$$\frac{1}{N}\sum_{i=1}^{N}\langle \vec{u}_i^2\rangle - \frac{1}{N_{r\,pairs}}\sum \langle \vec{u}_i \cdot \vec{u}_j\rangle \simeq \frac{1}{N}\sum_{i=1}^{N}\langle \vec{u}_i^2\rangle. \quad (11)$$

We are left with the high-frequency contribution to $\frac{1}{N}\Sigma_{i=1}^{N}\langle \vec{u}_i^2\rangle$ and it is reasonable to sum only over relevant modes and approximate

$$\frac{T(r)}{N} \simeq \frac{z\gamma}{3k_BT}\int_{\omega(r)}^{\omega(a)} \frac{3k_BT}{mN}\frac{g(\omega)}{\omega^2}d\omega = \frac{z}{N}\int_{l(r)}^{l(a)}\frac{g(l)}{l}dl. \quad (12)$$

Here $g(\omega)$ denotes the density of states, $\omega(a)$ and $\omega(r)$ are the frequencies associated with the mean spacing between nearest neighbors a and the distance r, respectively, m is the mass of a single bead in the elastic network, γ is the spring constant, $l \equiv m\omega^2/\gamma$ is a dimensionless variable, and the factor 3 appears due to the threefold degeneracy of the scalar elasticity model. In order to evaluate the integral we recall that $\omega(r) \sim r^{-d_f/d_s}$. On the other hand $\omega(a)$ is the basic frequency in the system and is hence of order $\sqrt{\gamma/m}$; it follows that $l(r) = C(r/a)^{-2d_f/d_s}$ where C is a numeric constant of order unity. The scaling of the density of states $g(l) \sim l^{d_s/2-1}$ is easily found by a change in variable when we recall that: $g(\omega) \sim \omega^{d_s-1}$. Preserving normalization, $\int_{l(2R_g)}^{l(a)} g(l)dl = N$, one gets

$$g(l) = \frac{Nd_s}{2C^{d_s/2}[1-(2R_g/a)^{-d_f}]}l^{d_s/2-1}, \quad (13)$$

where R_g is the radius of gyration. Substituting into Eq. (12) we integrate and obtain Eq. (1) where $\tilde{C} = \frac{zd_s}{2C[1-(2R_g/a)^{-d_f}]}$. We note that the expression we have obtained is clearly missing an additive constant since for a random walker traveling to a nearest neighbor $(r=a)$ it predicts zero MFPT.

Examining Eq. (1) more closely shows that the solution obtained via vibrational analysis complies with two basic requirements demanded from any solution of this problem: invariance under scaling and continuity with respect to the spectral dimension. Clearly if we were to take our RN and inflate/shrink it, creating an exact magnified/miniature copy of it, the MFPT between two sites will not be affected. This is so because the MFPT is only affected by the transition probabilities between sites and by construction these were left unchanged. It follows that $T(N,r)$ must be invariant under a scaling transformation that inflates/shrinks all lengths by a factor of α; Eq. (1) is clearly invariant under this transformation. Another thought experiment we could do is to think of a system whose spectral dimension is a tunable parameter. Examining the MFPT between two sites as we continuously vary the value of d_s, we would expect a continues behavior of $T(N,r)$. Indeed, taking the limit $d_s \to 2$ (from above/below) in Eq. (1) demonstrates the continuity of $T(N,r)$ with respect to the spectral dimension.

In this Rapid Communication we have introduced a vibrational shortcut to the solution of the MFPT problem on fractal structures and showed how the solution is readily obtained without the use of probabilistic arguments. Effective inverse spring constant interpretation of the MFPT allowed us to obtain the desired solution via scaling arguments. Direct calculation of the thermal variance in the distance between two tagged masses provides another route to the solution. Invariance under scaling and continuity with respect to the spectral dimension were shown to be emergent properties of the solution obtained via vibrational analysis. Our result emphasizes the duality between diffusion and vibrations on fractal structures. The study of diffusion and vibrations is essential to the understanding of biological systems where fractals were shown to naturally emerge. From the 3D structure and dynamics of single proteins [14–16] through the 3D organization of chromatin in the nucleus [17,18] and up to the entire cell level [19], fractals appear time and time again. As we have demonstrated above, when studying diffusion it is sometimes much more effective to tackle the dual problem of vibrations and vice versa. Even in the case where such a transformation does not carry with it an immediate computational gain, the insights gained from considering the analogous problem may be of interest. In Ref. [16] we discuss how such a procedure sheds new light regarding the fractal-like nature of proteins. In particular we utilize the inverse spring constant interpretation of the MFPT in the analysis of conformational changes in proteins.

S.R. acknowledges support from the Converging Technologies program of the Israeli Council for higher education. R.G. and J.K. acknowledge support from the Israel Science Foundation. R.G. acknowledges the Ilse Katz Center for Meso and Nanoscale Science and Technology and the Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering.

[1] B. B. Mandelbrot, *The (Mis)Behavior of Markets, A Fractal View of Risk, Ruin and Reward* (Basic Books, New York, 2004), pp. 186–195.

[2] S. Condamin, O. Benichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature (London) **450**, 77 (2007).

[3] B. D. Hughes and M. F. Shlesinger, J. Math. Phys. **23**, 1688 (1982).

[4] B. D. Hughes, *Random Walks: Random Walks and Random Environments* (Clarendon, Oxford, 1995),Vol. 2, Sec. 7.6.

[5] A. Telcs, *The Art of Random Walks* (Springer-Verlag, Vienna, 1885).

[6] T. Nakayama and K. Yakubo, *Fractal Concepts in Condensed Matter Physics*, Springer Series in Solid-State Sciences (Springer-Verlag, Berlin, 2003), p. 140.

[7] I. Bahar, A. R. Atilgan, and B. Erman, Folding. Des. **2**, 173 (1997).

[8] C. Chennubhotla and I. Bahar, PLOS Comput. Biol. **3**, e172 (2007).

[9] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, *Proceedings of the 21st ACM Symposium on the Theory of Computing* (ACM, New York, 1989), Vol. 574.
[10] A. P. Roberts and C. P. Haynes, http://arxiv.org/abs/0909.3353.
[11] S. Alexander, Phys. Rev. B **40**, 7953 (1989).
[12] Although it seems that our result in Eq. (11) depends on the approximation to follow, a more complete calculation that does not involve the neglect of the second term leads to effectively identical results.
[13] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[14] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008).
[15] M. de Leeuw, S. Reuveni, J. Klafter, and R. Granek, PLoS ONE **4**, e7296 (2009).
[16] S. Reuveni, R. Granek, and J. Klafter (unpublished).
[17] E. Lieberman Aiden *et al.*, Science **326**, 289 (2009).
[18] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. **23**, 373 (1993).
[19] A. Bancaud, EMBO J. **28**, 3785 (2009).

General mapping between random walks and thermal vibrations in elastic networks: Fractal networks as a case study

Shlomi Reuveni,[1,2] Rony Granek,[3] and Joseph Klafter[1]
[1]*School of Chemistry, Tel-Aviv University, Tel-Aviv 69978, Israel*
[2]*School of Mathematics, Tel-Aviv University, Tel-Aviv 69978, Israel*
[3]*The Stella and Avram Goren-Goldsten Department of Biotechnology Engineering, Ben-Gurion University, Beer Sheva 84105, Israel*
(Received 20 June 2010; revised manuscript received 12 September 2010; published 29 October 2010)

We present an approach to mapping between random walks and vibrational dynamics on general networks. Random walk occupation probabilities, first passage time distributions and passage probabilities between nodes are expressed in terms of thermal vibrational correlation functions. Recurrence is demonstrated equivalent to the Landau-Peierls instability. Fractal networks are analyzed as a case study. In particular, we show that the spectral dimension governs whether or not the first passage time distribution is well represented by its mean. We discuss relevance to universal features arising in protein vibrational dynamics.

DOI: 10.1103/PhysRevE.82.041132 PACS number(s): 05.40.Fb, 05.20.Gg, 05.45.Df

I. INTRODUCTION

Mapping two different physical problems onto one another has been proven very useful in physics. Examples are the mapping between the Schrodinger and diffusion equations [1], the mapping between lattice-gas and Ising models [2], the mapping between quantum field theories and critical phenomena [2], and the mapping between random walks and electric networks [3]. Here we focus on another well-known mapping, the mapping between random walks and vibrations in the scalar elasticity model. This mapping has been recognized for a long time, and some of its consequences have been already exploited, mainly in the context of fractal and percolation networks [4]. Nevertheless, despite the large amount of work done in the field, vibrational analogs for several basic quantities arising in the theory of random walks remained unknown.

Vibrations on a network of masses and springs are one of the most well studied problems in physics. Scalar elasticity is a special case where the three components of the displacement vector of a network node are decoupled from each other. Recently we have shown that the scaling behavior of the mean first passage time (MFPT) on fractal networks can be derived by exploiting an analogy with elastic networks [5]. The MFPT is an extremely important quantity governing the rate of diffusion controlled chemical reactions [6,7]. Chennubhotla *et al.* expressed the MFPT in terms of thermal vibrational correlation functions [8]. However, the applied line of argumentation could not be carried on to higher moments of the first passage time, nor to its full distribution or other probabilistic quantities. In the present study, we consider networks of general topology and focus on the construction of a rigorous mapping between the problem of a scalar elastic network coupled to a thermal bath and the random walk problem. The roots of our approach can be traced back to the pioneering work of Hattori *et al.* who established a relation between random walks and thermal correlation functions of spin systems [9].

II. MAPPING RANDOM WALKS TO THERMAL VIBRATIONS

Consider an elastic network (EN) of N masses coupled by harmonic springs in the framework of the scalar elasticity model also known as the Gaussian network model (GNM) when applied to proteins [4,10,11]. The GNM is defined by the quadratic Hamiltonian equation,

$$H_{GNM} = \sum_i \frac{m(\dot{\vec{u}}_i)^2}{2} + \frac{\gamma}{2}\sum_{i,j} \Delta_{ij}(\vec{u}_i - \vec{u}_j)^2. \tag{1}$$

The first term represents the kinetic energy of the system, γ is the spring force constant which is assumed to be homogeneous, $\vec{u}_i \equiv (x_i, y_i, z_i)$ and $\vec{R}_i = \vec{R}_i^0 + \vec{u}_i$ are the displacement with respect to the equilibrium position $\vec{R}_i^0$ and the instantaneous position of the ith mass, respectively. Δ is the network connectivity matrix with the following entries: $\Delta_{ij}=1$ if $i \neq j$ and the pair i,j is connected by a spring, $\Delta_{ij}=0$ otherwise. Here we will assume that a path of masses and springs exists between any two masses on the network. The GNM is threefold degenerate so it is sufficient to consider one spatial direction. Denoting by $\vec{x}$ the vector whose entries are x_i, the equations of motion in the absence of friction are: $m\frac{d^2\vec{x}}{dt^2} = -\gamma\Gamma\vec{x}$. Here Γ is the network Kirchhoff matrix,

$$\Gamma_{ij} = \begin{cases} -\Delta_{ij} & if \ \ i \neq j \\ z_i & if \ \ i = j \end{cases}, \tag{2}$$

where $z_i \equiv \Sigma_k \Delta_{ik}$ is the coordination number of the ith mass. Since rigid translations are a solution of the equations of motion, one of the eigenvalues of the matrix Γ is zero and consequently this matrix has no inverse. As an alternative, we exploit the existence of a generalized inverse. Denoting the Moore-Penrose pseudoinverse of Γ by Γ^{-1} and coupling the elastic network to a thermal bath one can show that [10,12]

$$\Gamma_{ij}^{-1} = \frac{\gamma}{k_B T}\langle x_i x_j \rangle_T. \tag{3}$$

where $\langle x_i x_j \rangle_T$ is the thermal correlation function.

The EN we have described above can also be thought of as a network of nodes connected by links (which we will refer to as RWN), see Fig. 1. We now construct a continuous

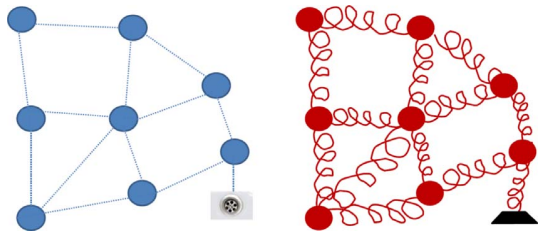


FIG. 1. (Color online) Left–a network of nodes connected by links (RWN), right—the corresponding elastic network of masses coupled by harmonic springs (EN). One can transform a RWN to an EN (and vice versa) by transforming nodes to masses and links to springs. Absorbing boundary conditions are tackled by introducing absorbing nodes and equivalent infinite mass walls. We study the relation between the random walk problem on the RWN and thermal vibrations of the EN.

time random walk on the RWN and show that under our construction the problem of vibrations on a network of masses and springs is practically equivalent to the random walk problem. In particular we express the *distribution of the first passage time* between two nodes (in Laplace plane) in terms of the vibrational correlation matrix defined in Eq. (3). Consider a random walk where the rate Γ_{jj} at which a random walker jumps out of node j is the coordination number of that node, $\Gamma_{jj}=z_j$, and that the random walker performs random jumps to its nearest neighbors without preference. In this scenario, $-\Gamma$ plays the role of a rate matrix whose elements $-\Gamma_{ij}$ are the transition rates from node j to node i. It follows that the change in occupation probabilities as a function of time follows the master equation: $\frac{d\vec{p}(t)}{dt}=-\Gamma\vec{p}(t)$, with the formal solution $\vec{p}(t)=\Sigma_{n=0}^{\infty}\frac{(-t\Gamma)^n}{n!}\vec{p}(0)\equiv\exp[-t\Gamma]\vec{p}(0)$.

When reflecting boundary conditions are introduced, probability is conserved. Indeed, from Eq. (2) it follows that: $\Sigma_{i\neq j}\Gamma_{ij}=-\Gamma_{jj}$, i.e., the rate of probability flow out of a site j is equal to the total rate at which probability flows into adjacent sites. The case of absorbing boundary conditions is tackled by introducing artificial absorbing nodes which are connected to some of the other network nodes. The EN analog is the introduction of equivalent infinite mass nodes which are connected by springs to the same masses as in the RWN. Mathematically this means that the network Kirchhoff matrix Γ would change such that $\Gamma_{ii}\to\Gamma_{ii}+n_i$, where n_i is the number of absorbing/infinite mass nodes connected to node i and we note that this would not change the form of the master equation. In addition, rigid translations are no longer a solution of the equations of motion. This implies that Γ has a true inverse Γ^{-1} for which Eq. (3) holds [12].

Denote by $P_{ij}(t)$ the probability that a random walker is found at site j at time t given that it was at site i at time $t=0$ and let $P(t)$ be the matrix whose entries are $P_{ij}(t)$. Denote by τ_{ij} the first passage time (FPT) (in the case of $i=j$ the first return time) of a random walker traveling from site i to site j and let $f_{ij}(t)$ be the probability density function of τ_{ij}. In the supplementary material accompanying this Letter we show that [12],

$$\begin{cases} \tilde{P}(s)=\dfrac{\hat{1}}{sN}\delta_{bc}+\dfrac{\Gamma^{-1}}{I+s\Gamma^{-1}} & \\ \tilde{f}_{ij}(s)=\dfrac{\dfrac{\delta_{bc}}{N}+\left(\dfrac{s\Gamma^{-1}}{I+s\Gamma^{-1}}\right)_{ij}}{\dfrac{\delta_{bc}}{N}+\left(\dfrac{s\Gamma^{-1}}{I+s\Gamma^{-1}}\right)_{jj}} & i\neq j \\ \tilde{f}_{ii}(s)=1-\dfrac{s}{(s+z_i)\left[\dfrac{\delta_{bc}}{N}+\left(\dfrac{s\Gamma^{-1}}{I+s\Gamma^{-1}}\right)_{ii}\right]} & i=j. \end{cases} \tag{4}$$

Here, $\tilde{P}(s)$, $\tilde{f}_{ij}(s)$, $\tilde{f}_{ii}(s)$ are the Laplace transforms of $P(t)$, $f_{ij}(t)$, $f_{ii}(t)$, respectively, $\hat{1}$ is a matrix whose *all* entries equal one and δ_{bc} equals either one or zero in the case of reflecting/absorbing boundary conditions, respectively. Equation (4) is the fundamental result that stands in the basis of this paper. It provides the connection between the statistical mechanics of an elastic network and the stochastic dynamics of a random walker on the same network. $P(t)$ and $f_{ij}(t)$ are uniquely determined by their Laplace transform and it follows that all the information about the stochastic dynamics of a random walker is contained in the pairwise correlation matrix Γ^{-1}. We proceed with the derivation of novel relations implied by Eq. (4).

III. ABSORBING BOUNDARY CONDITIONS

On a finite network with reflecting boundary conditions the probability of a random walker to eventually reach a target site j given it started at i is one. In general this probability is given by $F_{ij}=\int_0^\infty f_{ij}(t)dt=\lim_{s\to 0}\tilde{f}_{ij}(s)$ and our conclusion is revalidated by taking this limit in Eq. (4) ($\delta_{bc}=1$). The case of absorbing boundary conditions is more interesting since the random walker may be absorbed at the boundary before reaching its destiny. Taking the limit $s\to 0$ ($\delta_{bc}=0$) in Eq. (4) we obtain a vibrational interpretation for F_{ij} on a finite network equipped with absorbing boundary conditions,

$$\begin{cases} F_{ij}=\dfrac{\Gamma_{ij}^{-1}}{\Gamma_{jj}^{-1}}=\dfrac{\langle x_i x_j\rangle_T}{\langle x_j^2\rangle_T} & i\neq j \\ F_{ii}=1-\dfrac{1}{z_i\Gamma_{ii}^{-1}}=1-\dfrac{1}{\dfrac{\gamma z_i}{k_BT}\langle x_i^2\rangle_T} & i=j. \end{cases} \tag{5}$$

Doing the same for $\tilde{P}(s)$ yields a vibrational interpretation for the mean time spent in site j (given the walk started at site i) prior to absorption: $\int_0^\infty P_{ij}(t)dt=\frac{\gamma}{k_BT}\langle x_i x_j\rangle_T$. Interestingly, F_{ij} depends only upon the vibrational correlation between source and target and the vibrational mean square displacement (MSD) of the target. Similarly, the probability to eventually return to the origin F_{ii}, depends only upon the vibrational MSD of the origin and its coordination number z_i. We note that by definition $1\geq F_{ij}$, $F_{ii}\geq 0$ and hence in the case of absorbing boundary conditions: $\langle x_i^2\rangle_T\geq\frac{k_BT}{\gamma z_i}$ and $\langle x_j^2\rangle_T\geq\langle x_i x_j\rangle_T\geq 0$. In the scalar elasticity model the vector $\vec{x}$

is multivariate normal [10] and its equilibrium distribution is uniquely determined by the correlation matrix $\langle x_i x_j \rangle_T$. Since $\int_0^\infty P_{ij}(t)dt$ determines $\langle x_i x_j \rangle_T$, the stochastic dynamics of a random walker on the RWN uniquely determines the equilibrium distribution of the displacements vector on the EN.

Equation (5) allows us to examine the question of recurrence in infinite networks by taking the thermodynamic limit. In achieving this limit, we construct an infinite network as a limit of on growing finite subnetworks. Doing so, we keep all absorbing nodes at the periphery as subnetworks grow in size. We discriminate between two types of infinite elastic networks, those that are thermodynamically stable and those that are not. From Eq. (5) it is clear that $\lim_{N\to\infty} F_{ii}=1 \Leftrightarrow \lim_{N\to\infty}\langle x_i^2\rangle_T=\infty$ and that $\lim_{N\to\infty} F_{ii}<1 \Leftrightarrow \lim_{N\to\infty}\langle x_i^2\rangle_T<\infty$. A network is called recurrent if $F_{ii}=1$ for every site i [13,14]. By assumption there is path connecting any two nodes and one can show that in this case F_{ii} are either all equal to one or all smaller than one [14]. Correspondingly $\langle x_i^2\rangle_T$ are either all infinite or all finite and we conclude that an infinite network is recurrent if it is thermodynamically unstable and vice versa. Recurrence, in the context of random walks, is hence a term equivalent to the Landau-Peierls instability of elastic networks [15–17]. Note that no assumptions regarding network structure were made in arriving to this conclusion.

Since $\langle x_i^2\rangle_T$ either all diverge or all converge in the thermodynamic limit, considering the average vibrational MSD: $\langle x^2\rangle_T=\frac{1}{N}\Sigma_{n=1}^N\langle x_i^2\rangle_T$ is usually enough in order to determine if an infinite network is recurrent. On fractal elastic networks the evaluation of $\langle x^2\rangle_T$ is relatively simple since after going to normal modes and invoking the equipartition theorem one is left with a simple integral that gives [16]

$$\langle x^2\rangle_T \sim \begin{cases} \text{constant} & d_s > 2 \\ ln(N) & d_s = 2 \\ N^{(2/d_s)-1} & d_s < 2. \end{cases} \tag{6}$$

Here d_s is the network spectral dimension that governs the density of vibrational modes $g(\omega)$ at low frequencies via the scaling relation $g(\omega)\sim\omega^{d_s-1}$. For regular three-dimensional (3D), two-dimensional (2D), and one-dimensional (1D) networks the spectral dimension coincides with the regular dimension and one recovers the well known Debye density of states. Combing Eqs. (5) and (6) we revalidate that an infinite fractal network is recurrent if $d_s \leq 2$ (and vice versa) a result which was first obtained by Burioni *et al.* [16]. In addition we recover the average manner in which the limit $F_{ii}\to 1$ for $N\to\infty$ is achieved. Note that one must be more careful applying the above arguments when considering pathological cases for which, after taking the thermodynamic limit, $\langle x^2\rangle_T$ is infinite despite the fact that $\langle x_i^2\rangle_T$ are all finite and we refer the reader to a discussion on the related issue of "recurrence on the average" [14].

IV. REFLECTING BOUNDARY CONDITIONS

On a finite network with absorbing boundary conditions for every two sites characterized by $F_{ij}<1$, the probability that the random walker would never reach j ($\tau_{ij}=\infty$) is positive and the MFPT is therefore infinite. In contrast, in the case of reflecting boundary conditions, the random walker is sure to reach its destiny and the MFPT is then the single most important quantity describing τ_{ij}. The kth moment of τ_{ij} is given by: $E[\tau_{ij}^k]=(-1)^k\tilde{f}_{ij}(s)^{(k)}|_{s=0}$. Taking the first derivative and the limit $s\to 0$ ($\delta_{bc}=1$) in Eq. (4) we obtain

$$E[\tau_{ij}] = \begin{cases} \frac{N\gamma}{k_B T}[\langle x_j^2\rangle_T - \langle x_i x_j\rangle_T] & i \neq j \\ \frac{N}{z_i} & i = j. \end{cases} \tag{7}$$

Interestingly, the MFPT from site i to site $j\neq i$ depends only upon the vibrational correlation between source and target and the vibrational MSD of the target. The mean first return time is only affected by the number of nodes N and the coordination number z_i. Note that $E[\tau_{ij}]>0$ by definition and hence: $\langle x_j^2\rangle_T > \langle x_i x_j\rangle_T$, in contrast to the case of absorbing boundary conditions $\langle x_i x_j\rangle$ can be negative.

Equation (7) can be used in order to obtain a simple vibrational derivation of the scaling law for the global MFPT (GMFPT) on a finite fractal domain [18]. The global first passage time (GFPT) is defined as the random time it takes a random walker to reach a randomly selected target from a randomly selected origin. Here we assume that the source and target are different sites. The GMFPT is defined as the average GFPT: $\text{GMFPT}(N)=\frac{1}{N(N-1)}\Sigma_{i,j,i\neq j}E[\tau_{ij}]$. Using Eq. (7) for $E[\tau_{ij}]$ and assuming that cross correlations add up (incoherently) to a negligible contribution, we obtain: $\text{GMFPT}(N)\simeq\frac{N\gamma}{k_b T}\langle x^2\rangle_T$. The scaling of the GMFPT with the number of nodes follows from Eq. (6),

$$\text{GMFPT}(N) \sim \begin{cases} N & d_s > 2 \\ N\, ln(N) & d_s = 2 \\ N^{2/d_s} & d_s < 2. \end{cases} \tag{8}$$

The MFPT does not provide a complete characterization of the FPT distribution. Another important, and much less studied quantity, is the variance. The variance gives a measure for the width of the FPT distribution around the mean. If the width is very large in comparison to the MFPT, the latter cannot be considered a reliable representative of the distribution. For $i=j$ we use Eq. (4) ($\delta_{bc}=1$) to obtain [12]: $E[\tau_{ii}^2]=\tilde{f}_{ij}(s)^{(2)}|_{s=0}=2(\frac{N}{z_i})^2[\frac{1}{N}+\frac{z_i\gamma}{k_b T}\langle x_i^2\rangle_T]$ and for the reduced variance [12]: $\sigma^2[\tau_{ii}]/E^2[\tau_{ii}]=\frac{2}{N}+\frac{2z_i\gamma}{k_B T}\langle x_i^2\rangle-1$. In fractal networks with a coordination number that is sharply distributed around a mean value z, we can replace z_i with z, average over all sites and use Eq. (6) to obtain

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2[\tau_{ii}]}{E^2[\tau_{ii}]} \sim \begin{cases} \text{const} & d_s > 2 \\ ln(N) & d_s = 2 \\ N^{2/d_s-1} & d_s < 2. \end{cases} \tag{9}$$

We conclude that in the case of $d_s>2$ the standard deviation is of the same order of the MFPT regardless of the number of nodes N. Conversely, when $d_s<2$ the reduced variance diverges as a power law of N and the MFPT cannot be considered a reliable representative of the FPT distribution. When

$i \neq j$ we average over all pairs distanced r apart and obtain the following approximation [12]:

$$\frac{1}{N_r}\sum_{i,j}\frac{\sigma^2[\tau_{ij}]}{E^2[\tau_{ij}]} \sim \begin{cases} [1-(r/a)^{d_f(2/d_s-1)}]^{-1} & d_s > 2 \\ ln(N)/ln(r/a) & d_s = 2 \\ N^{2/d_s-1}(r/a)^{-d_f(2/d_s-1)} & d_s < 2. \end{cases} \quad (10)$$

Here N_r is the number of pairs distanced r apart and the sum goes only over these pairs, a is the average distance between nearest neighbors and d_f is the network fractal dimension. We note that the joint domain in which the approximation is valid is given by: $\{N \gg 1, a \ll r \ll R_g\}$ where R_g is the radius of gyration. Equation (10) demonstrates that the reduced variance is a monotonically decreasing function of the distance between source and target. Keeping this distance fixed, the reduced variance is independent of the number of nodes in the case of $d_s > 2$ but when $d_s \leq 2$ it diverges with the number of nodes. Equation (10) stands in line with a similar result obtained independently by Bénichou *et al.* [7].

V. SUMMARY AND DISCUSSION

In this paper, we have rigorously mapped observables that appear in random walk theory to observables associated with the problem of an elastic network coupled to a thermal bath. Our mapping provides a systematic way for translating random walk problems to the realm of elasticity physics allowing for new theoretical, computational, and experimental approaches toward the random walk/elastic network problems. The mapping was demonstrated useful in the analysis of random walk problems on complex, scale invariant, media. It is important to emphasize that the MFPT heavily depends on network topology. The entire network topology is summarized in the network Kirchhoff matrix Γ and hence [using Eq. (7)] knowing the network topology amounts to knowing the MFPT. More surprising is the converse statement. Network topology can be reconstructed knowing the MFPT from any node to any other node and the second moment of the first return time [12].

Recently, we have utilized random walks on protein structures to study the vibrational dynamics of proteins [19]. In particular, we used a special case of Eq. (7) [8] relating the MFPT with the thermal variance in the instantaneous distance between amino acids. This equation has been used to unravel universal properties in the vibrational dynamics of proteins. In addition, we have shown that a sharp deviation from the mean universal behavior may result in the emergence of specific functionality. We suggest that exploiting the different relations between vibrations and random walks, as stated in this paper, can be beneficial in the research of functionality-dynamics interplay in proteins.

ACKNOWLEDGMENTS

S.R. acknowledges support from the Converging Technologies program of the Israeli Council for higher education. R.G. and J.K. acknowledge support from the Israel Science Foundation and the Deutsch-Israelische Projektkooperation (DIP). R.G. acknowledges support from the Ilse Katz Center for Meso and Nanoscale Science and Technology and the Reimund Stadler Minerva Center for Mesoscale Macromolecular Engineering.

[1] L. S. Schulman, Techniques and Applications of Path Integration Dover Books on Physics (Dover Publications, New York, 2005).
[2] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, England, 2007).
[3] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, Carus Mathematical Monographs (Mathematical Association of America, Washington, 1984).
[4] T. Nakayama, K. Yakubo, *Fractal Concepts in Condensed Matter Physics*, Springer Series in Solid-State Sciences Vol. 140 (Springer-Verlag, Berlin, 2003).
[5] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. E **81**, 040103(R) (2010).
[6] S. Condamin, O. Benichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature (London) **450**, 77 (2007).
[7] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer, and R. Voituriez, Nat. Chem. **2**, 472 (2010).
[8] C. Chennubhotla and I. Bahar, PLOS Comput. Biol. **3**, e172 (2007).
[9] K. Hattori, T. Hattori, and H. Watanabe, Prog. Theor. Phys. **92**, 108 (1987).
[10] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[11] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[12] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevE.82.041132 derivations of equations (3), (4), (7), (10) and supplementary discussion.
[13] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, New York, 2001).
[14] R. Burioni and D. Cassi, J. Phys. A **38**, R45 (2005).
[15] R. Peierls, Helv. Phys. Acta **81**, 2 (1934).
[16] R. Burioni, D. Cassi, M. P. Fontana, and A. Vulpiani, EPL **58**, 806 (2002).
[17] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008).
[18] V. Tejedor, O. Bénichou, and R. Voituriez, Phys. Rev. E **80**, 065104(R) (2009).
[19] S. Reuveni, R. Granek, and J. Klafter, Proc. Natl. Acad. Sci. U.S.A. **107**, 13696-13700 (2010).

10 Summary and Discussion

A recent review article dedicated to the fractal-like nature of proteins opens with the following paragraph [85]:

"The application of fractal dimension-based constructs to probe the protein interior dates back to the development of the concept of fractal dimension itself. Numerous approaches have been tried and tested over a course of (almost) 30 years with the aim of elucidating the various facets of symmetry of self-similarity prevalent in the protein interior. In the last 5 years especially, there has been a startling upsurge of research that innovatively stretches the limits of fractal-based studies to present an array of unexpected results on the biophysical properties of protein interior."

Indeed, in the last few years we are experiencing a fresh wave on renewed interest in the fractal point of view on protein structure and dynamics. Moreover, as newly discovered theoretical predictions motivate novel experimental work and vice versa, this wave is not soon to go away. In this chapter, I will describe our own contributions to the field, summarizing and recapitulating the highlights of our studies while putting a special emphasis on novel aspects of our work.

10.1 Foundations

Experimental evidence of fractality in proteins came from electron spin relaxation measurements [50] and neutron scattering [51]. Indirect evidence came from single molecule experiments that have reported anomalous behaviors involving power-laws in time [86, 87], from neutron spin-echo studies that measured a stretched exponential decay of the dynamic structure factor [71] and from molecular dynamics simulations that have shown anomalous diffusion of vibrational energy [53], dihedral angels [88] and amino acids [89].

A conceptual framework began to emerge when, in analogy with mathematically constructed fractals, it became clear that each protein can be associated with unique characteristic broken dimensions. The first seeds were planted in the work of Burioni *et al.* [33] who computed the spectral dimension for a set of 57 proteins and the work of Enright *et al.* [8] who computed the fractal dimension for a set of 200 proteins. We have unified and significantly elaborated the analysis made by the above mentioned groups in two sequential stages. In the first stage we created a joint data base of 543 proteins for which we have computed both the fractal and spectral dimensions [47]. We have then acted to automate the procedure according to which these dimensions are calculated. After succeeding in doing so, we created a joint data base of more than 5000 proteins for which the fractal and spectral dimensions were calculated and are now freely available on the web [48].

The fractal-like nature of proteins is manifested in the fact that the fractal and spectral dimensions associated with them, are both broken dimensions

which are lower than the (integer) dimension of the embedding space. Proteins belong to realm of low dimensionality which gives rise to a sparser fill of space ($d_f < 3$) and increased flexibility due to an excess of low frequency vibrational modes and the effect of the Landau-Peierls instability ($d_s < 2$).

An attempt to tie together the spectral dimension, d_s, and the number of amino acids, N, have been made in the past [33]. In similarity to our approach, this attempt was also based on the Landau-Peierls instability criterion. However, the agreement between the theoretical prediction and observations based on analysis of protein structures obtained via X-ray crystallography was unsatisfactory. As we have later found, a relation between the spectral dimension and the number of amino acids does in fact exist. A delicate balance between mutual requirements for stability and flexibility forces proteins to obey an 'equation of state' in the fractal parameter space [47, 48]:

$$\frac{2}{d_s} + \frac{1}{d_f} = 1 + \frac{b}{ln(N)} \tag{77}$$

One can easily note the appearance of the fractal dimension on the left hand side of Eq. 77. Indeed, in order to derive it, the important contribution of the fractal dimension, i.e. the sparse 'space filling' property of proteins, must be taken into account. The indispensability and crucial role played by a joint data base in which both the fractal and spectral dimension are readily available is thus clear.

Equation 77 should not be taken light-headedly. As each and every protein is, in fact, a different physical system and in light of the enormous diversity among proteins, the mere existence of a universal law obeyed by all proteins alike is extremely surprising. And yet, as we have demonstrated, a unifying, quantifiable, common denominator does exist. Adopting a fractal point of view on proteins, we were able to unravel universality in a realm of specificity.

Striving for additional, independent, support for the fractal-like nature of proteins we have introduced a random walk approach to the study of protein structure and dynamics [49]. As indicated in the table below, we have noted two strong links between vibrational dynamics of fractal networks and random walks on them [41, 59]:

Random Walks	Vibrational Dynamics
$P_0(t) \sim t^{-d_s/2}$	$g(\omega) \sim \omega^{d_s-1}$
$MSD(t) \sim t^{2/d_w}$	$< \triangle^2(t) > \sim t^{2/d_w}$

The density of vibrational states, $g(\omega)$, and the probability of a random walker to be found at the origin at time t, $P_0(t)$, are both characterized by the spectral dimension. The propagation of vibrational energy on a fractal network and the mean square displacement of random walker on it, are related as well. In the overdamped limit, the mean square propagation length of an initially localized structural deformation, $< \triangle^2(t) >$, and the mean square displacement of a random walker, $MSD(t)$, both scale as $\sim t^{2/d_w}$. This scaling relation serves as the definition of the walk dimension d_w.

It is thus possible to obtain an independent assessment of the spectral dimension of proteins via random walk (or equivalently diffusion) simulations. Similar simulations can also teach us how vibrational energy spread on the protein fold. Here again we have found a striking difference between proteins and regular 3D structures [49]. While for the latter the spectral dimension coincides with the embedding dimension ($d_s = 3$), the normal $P_0(t) \sim t^{-3/2}$ law is recovered and regular diffusion is found $MSD(t) \sim t$ ($d_w = 2$); proteins are characterized by $d_s < 2 < 3$, an anomalously slow decay of $P_0(t)$ and anomalous subdiffusion for which the mean square displacement grows sublinearly with time ($d_w > 2$).

Subdiffusion is a characteristic property of random motion on fractal structures. Hence, our observations provide additional support for the fractal-like nature of proteins. The mapping between random walks and vibrational dynamics asserts that the above mentioned findings are also an independent revalidation of vibrational studies which have demonstrated non-Debye density of vibrational states [33, 47, 48, 50, 51] and anomalous subdiffusion of vibrational energy [52, 53]. And so, random walk simulations and vibrational dynamics studies, being different facets of the same phenomenon, are both in accord with the fractal-like nature of proteins.

Foundations - Summary

- **Thousands of proteins were studied using a variety of methods and were found to be fractal-like.**
- **The spectral dimension of proteins is smaller than two ($d_s < 2$), a fact which gives rise to increased flexibility.**
- **The fractal dimension of proteins is smaller than three ($d_f < 3$), a fact which gives rise to a sparser fill of space.**
- **The walk dimension of proteins is larger than two ($d_w > 2$), a fact which gives rise to anomalous subdiffusion of vibrational energy.**
- **A delicate balance between stability and flexibility forces proteins to obey an 'equation of state' in the fractal parameter space:**

$$\frac{2}{d_s} + \frac{1}{d_f} = 1 + \frac{b}{ln(N)}$$

10.2 Applications

After having established the foundations of the fractal approach we turned to find further implications and applications of it. We were once more motivated by what seemed to be a contradiction. Vibrating under the influence of a complex energy landscape, proteins display rich and intricate dynamics. However, near its minima, the energy landscape has a simple harmonic form that does not seem to coincide with the above mentioned behavior. It is therefore not at all clear whether simple harmonic models, such as the GNM, are able to reproduce the observed complexity in the vibrational dynamics of proteins.

In light of the above, one of the most important conclusions of our study is the fact that the fractal-like structure of proteins is sufficient in order to explain anomalous vibrational dynamics [49]. The native state structure of proteins is anomalous in its basis. As a result one needs not go beyond the harmonic approximation in order to observe anomalous dynamics since anomalous structure leads to it, even in the sole presence of harmonic interactions. Interestingly, when the structure is fractal-like, large conformational changes are encoded within the native state structure and are implied by the harmonic approximation rather than lead to its breakdown. This conclusion is in fact an additional facet of the Landau-Peierls instability and yet it is rather illuminating to view it in light of the mapping between random walks and vibrational dynamics.

The thermal variance, $< (\triangle R_{ij})^2 >$, in the distance, R_{ij}, between two amino acids is an experimentally measurable evidence of protein flexibility [90]. This variance is a result of thermal fluctuations and its magnitude is governed by the mechanical properties of the protein. Chennubhotla *et al.* were the first to realize that, within the framework of the GNM, the mean time it takes a random walker to get from one amino acid to the other is equivalent to the thermal variance in the distance between the same two amino acids [58]. Interestingly, on the exact same year the mean first passage time problem was solved for fractal networks by Condamin *et al.* [81]. Being aware of the mutual existence of these novel discoveries we were in a position that allowed us to translate results originally obtained for the mean first passage time problem to a vibrational language.

For fractal networks characterized by a spectral dimension lower than two the mean first passage time was shown to scale as a power law of the distance between the source and target nodes. A similar result regarding the average variance in the distance between two beads on a fractal elastic network immediately follows. Applied to proteins, this result predicts a power law divergence of the thermal variance in the distance between amino acids with respect to the equilibrium distance between them: $< (\triangle R_{ij})^2 > \sim (R_{ij}^0)^{\alpha}$. The steep dependence of variance upon distance allows the existence of large conformational changes. It is important to note that this dependence is significantly different from the dependence observed for elastic networks characterized by a spectral dimension larger than two. In these networks distances between beads vary within a relatively small range and large conformational changes are improbable.

Anomalous structure can also lead to an anomalous, non exponential, decay of the dynamic structure factor. The dynamic structure factor is measured via scattering experiments. The basic setup and theory that stand behind these experiments is probably known to every undergraduate physics student worldwide. One of the classical results in this field goes to the decay of the static structure factor with the wavenumber k, $S(k) \sim k^{-d}$, where d is the dimensionality of the lattice. This result is further generalized in various textbooks to the case of a fractal networks for which the integer dimension, d, is replaced by the fractal dimension d_f [36, 37]. The fractal dimension is hence experimentally measurable via structure factor studies.

While the static structure factor of fractals is well understood, our understanding of the dynamic structure factor is much more limited. In the context of solid fractals, the dynamic structure factor has been extensively analyzed on the 'single phonon' level, and in the absence of any source of friction. This provides a good description for the inelastic (Brillouin) scattering from solid fractals, but is not adequate for the quasi-elastic scattering from low dimensional fractals in solutions due to large fluctuations and friction dominated dynamics.

Neutron spin-echo studies that measure the dynamic structure factor $S(k,t)$ have been recently performed on horse heart myoglobin and bovine hemoglobin in solutions [71]. In the large wavenumber k regime corresponding to $kR_g \gg 1$, where R_g is the gyration radius, and at low concentrations and times shorter than 1ns, the result is a stretched exponential relaxation, $S(k,t) \sim e^{-(\Gamma_k t)^\nu}$ with $\nu \simeq 0.4 \pm 0.03$ for both proteins and independent of k. The relaxation rate has been found to scale as $\Gamma_k \sim k^{2/\nu}$. Inelastic neutron scattering, a complementary method to neutron spin-echo, is used for exploring protein dynamics at high frequencies ('short times') [91, 92]. Inelastic neutron scattering experiments and molecular dynamics simulations performed on lysozyme showed a non-Lorentzian spectra which corresponds to a non-exponential decay of the dynamic structure factor [93]. These findings are in accord with the neutron spin-echo findings described above.

Phenomenologically, the non-exponential relaxation of $S(k,t)$ can be approached by assuming that collective (averaged) observables, associated with the internal dynamics of proteins, follow fractional Brownian motion rather than regular Brownian motion [94, 95]. The assumption is justified a posteriori by comparison to molecular dynamics simulations and experiments. Another approach towards the anomalous relaxation of $S(k,t)$ is based on an analogy with polymer dynamics theory and one conjectures that $S(k,t)$ decays mainly due to the time evolution of the amino acid mean square displacement [37, 62, 96, 97]. Combined with the above experimental observations, this conjecture suggests that the mean square displacement evolves anomalously in time, i.e., as $\sim t^\nu$ with $\nu < 1$. Yet, the anomalous diffusion exponent does not fit any of the polymer theory exponents, 2/3 in the Zimm model and 1/2 in the Rouse model. More importantly, a folded protein is clearly nothing like a solvated, open, flexible polymer, that fluctuates between its many available configurations without having any underlying scaffold. Hence, the use of polymer theory can only serve as a guiding tool.

We have taken a first principles approach in computing the decay of the dynamic structure factor of fractal elastic networks in general and proteins in particular [68, 72]. Since all amino acids contribute to the dynamic structure factor, it should indeed be strongly influenced by the spatially averaged mean square displacement. To this end, we have first calculated the time evolution of the latter and showed that due to the protein fractal-like structure, the mean square displacement of an amino acid, averaged over all amino acids of the protein, scales at short times as $\sim t^{\nu}$. The exponent ν depends on the fractal and spectral dimensions, $\nu = 1 - d_s/2$ in a Rouse-type model where the friction is local [57, 49, 96, 97], $\nu = (2-d_s)/(2-d_s+d_s/d_f)$ in a Zimm-type model where friction arises from long ranged hydrodynamic interactions [98], and $\nu = 2 - d_s$ for vanishing friction [57].

However, as previously mentioned, a mean square displacement based picture is an averaged one, and as such a bit oversimplified and incomplete. Such a picture fails, for example, to capture both static and dynamic correlations between amino acids, e.g., it is unable to predict the static structure factor $S(k)$. We have thus deepened our analysis beyond the averaged mean square displacement picture. Extending the well known, 'text book', result for $S(k)$ from the spatial regime to the spatial-temporal regime, we have made a complete calculation of the dynamic structure factor delineating the regimes for which the decay of $S(k,t)$ is dominated by the mean square displacement.

Our main result regarding the decay of the dynamic structure factor of fractals can be simply stated as follows. Assume a large bead-spring fractal network with $d_s < 2$ and arbitrary d_f, and consider a scattering experiment at large wavenumbers k such that both $kR_g \gg 1$ and $k\bar{u} \gg 1$, where $\bar{u} \equiv \sqrt{\langle u^2 \rangle}$ is the spatially averaged root mean square displacement. In this limit, and within a wide window of time, we find that $S(k,t)$ decays as a stretched exponential $S(k,t) \simeq S(k)e^{-(\Gamma_k t)^{\nu}}$, where the relaxation rate anomalously depends on k, $\Gamma_k \sim k^{2/\nu}$ and ν is the anomalous diffusion exponent mentioned above. Our result generalizes known results for the linear Rouse polymer chain ($d_s = 1$), and for the linear-Gaussian Zimm polymer chain ($d_s = 1$, $d_f = 2$) [37]. Moreover, our result relate the fractal and spectral dimensions d_s and d_f to the stretched exponential decay of $S(k,t)$ and allow for experimental measurement of the these dimensions.

Besides proteins, our theory can be used to predict the decay of the dynamic structure factor in various other systems as well: (i) Glass forming colloidal suspensions [99]. Assuming analogy to 3D percolation network, that implies $d_f \simeq 2.48$ and $d_s \simeq 1.328$, and accounting for the hydrodynamic coupling that yields $\nu = (2 - d_s)/(2 - d_s + d_s/d_f)$, we suggest a stretched exponential decay with $\nu \simeq 0.556$, remarkably close to the observed value $\nu \simeq 0.6$; (ii) Colloidal gels [100], that show a clear fractal structure and for which a Zimm-type dynamics and bond-bending potential explains well the observed stretched exponential decay [98]; and (iii) Chromatin [65, 66, 101, 102, 103], for which it was recently shown that telomeres perform anomalous subdiffusion with $\nu \simeq 0.32$ [104]. This may be interpreted, within the Rouse model that yields $\nu = 1 - d_s/2$ (noting

that in such a dense polymer system hydrodynamics is likely to be screened), by $d_s \simeq 1.36$. This value of d_s is remarkably close to that of percolation clusters in $2 < d < 5$ dimensions and suggests the presence of DNA crosslinks (e.g., via ligation). According to the present calculation, it is suggested that the DSF of chromatin will decay as a stretched exponential with stretching exponent $\simeq 0.32$, which can motivate experiments in this direction.

Applications - Summary

- **Fractal-like structure is the origin of anomalous vibrational dynamics in proteins.**
- **The spatially averaged thermal variance in the distance between two network beads (amino acids) is predicted to scale as a power law of the equilibrium distance between them:** $< (\triangle R_{ij})^2 > \sim (R_{ij}^0)^\alpha$.
- **The time dependent vibrational mean square displacement of a network bead (amino acid) is predicted to be subdiffusive:** $< \left(\vec{u}_i(t) - \vec{u}_i(0)\right)^2 > \sim t^\nu$ **with** $\nu < 1$.
- **The decay of the dynamic structure factor,** $S(k,t)$**, is dominated by the spatially averaged mean square displacement of a network bead (amino acid). As a result,** $S(k,t)$ **decays as a stretched exponential** $S(k,t) \approx S(k) e^{-(\Gamma_k t)^\nu}$ **with** $\Gamma_k \sim k^{2/\nu}$.
- **The anomalous diffusion and stretching exponent,** ν**, depends on the fractal and spectral dimensions,** $\nu = 1 - d_s/2$ **in a Rouse-type model where the friction is local,** $\nu = (2 - d_s)/(2 - d_s + d_s/d_f)$ **in a Zimm-type model where friction arises from long ranged hydrodynamic interactions, and** $\nu = 2 - d_s$ **for vanishing friction.**

10.3 Ramifications

The mapping between random walks and vibrations has long been recognized and some of its consequences have already been exploited, mainly in the context of fractal and percolation networks [41, 59]. More recently, while studying proteins, we have utilized this mapping as well. It is nevertheless true that despite the large amount of work done in the field, vibrational analogs for several basic quantities arising in the theory of random walks remained unknown for a long period of time. In particular, a connection between the first passage time problem and vibrational dynamics was not known.

Chennubhotla *et al.* found that the thermal variance in the distance between two network beads is the vibrational equivalent of the mean first passage time [58]. Following this seminal discovery we have revisited the mapping between random walks and thermal vibrations and further advanced upon it [83, 84]. Our main findings are summarized in the box below.

Ramifications - Summary

- **The mean first passage time problem has a mechanical analog that allows a shortcut towards its solution on self similar networks. The solution is readily obtained using basic, bead spring, physics and scaling arguments.**
- **Random walk occupation probabilities, first passage time distributions and passage probabilities between nodes can all be interpreted in terms of thermal vibrational correlation functions.**
- **A random walker will return to the origin with probability one, if and only if the walk is performed on a network prone to Landau-Peierls instability.**
- **Mean first passage time based arguments must be used carefully since fluctuation around the mean may dominant the first passage time distribution. As long as the spectral dimension is smaller than two, the standard deviation is of the same order magnitude as the mean, regardless of the size of the network. However, when the spectral dimension is smaller than two, the ratio between the standard deviation and the mean diverges as a power law of number of network nodes. In these cases the mean first passage time is clearly insufficient and one must take into account higher moments of the first passage distribution.**

References

[1] J. B. Sumner, The Isolation and Crystallization of the Enzyme Urease. *J. Biol. Chem.,* **69**, 435, (1926).

[2] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W.H. Freeman, (2004).

[3] H. Muirhead and M. Perutz, Structure of hemoglobin, a three-dimensional Fourier synthesis of reduced human hemoglobin at 5.5A resolution. *Nature,* **199**, 633, (1963).

[4] J. Kendrew, G. Bodo, H. Dintzis, R. Parrish, H. Wyckoff and D. Phillips, A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* **181**, 662, (1958).

[5] F. C. Bernstein *et al.*, The protein data bank: a computer-based archival file for macro-molecular structure. *J. Mol. Biol.,* **112**, 535, (1977).

[6] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman, (1977).

[7] K. Falconer, *Fractal Geometry: Mathematical Foundations and Application,* John Wiley & Sons, (2003).

[8] M. B. Enright and D. M. Leitner, Mass fractal dimension and the compactness of proteins. *Phys. Rev. E,* **71**, 011912, (2005).

[9] D. Whitford*, Proteins: Structure and Function,* John Wiley & Sons, (2005).

[10] K. Tai, Conformational sampling for the impatient. *Biophys. Chem.*, **107**, 213, (2004).

[11] J. Norberg and L. Nilsson, Advances in biomolecular simulations: methodology and recent applications. *Q. Rev. Biophys.,* **36**, 257, (2003).

[12] M. Karplus, Molecular dynamics of biological macromolecules: a brief history and perspective. *Biopolymers,* **68**, 350, (2003).

[13] R. Schleif, Modeling and studying proteins with molecular dynamics. *Methods Enzymol,* **383**, 28, (2004).

[14] J. A McCammon, B. R. Gelin and M. Karplus, Dynamics of folded proteins. *Nature,* **267**, 585, (1977).

[15] M. Karplus and J. Kuriyan, Molecular dynamics and protein function. *Proc. Natl. Acad. Sci.,* **102**, 6679, (2005).

[16] V. Tozzini, Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.,* **15**, 144, (2005).

[17] Q. Cui (Editor) and I. Bahar (Editor), *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems,* Chapman & Hall/CRC, (2005).

[18] J. Ma, Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, **13**, 373, (2005).

[19] I. Bahar and A. J. Rader, Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **15**(5), 586, (2005).

[20] M. M., Tirion, Large amplitude elastic motions in proteins from a single-parameter atomic analysis, *Phys. Rev. Lett.,* **77**, 1905, (1996).

[21] I. Bahar, A. R. Atilgan and B. Erman, Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Fold. Des.,* **2**, 173, (1997).

[22] T. Haliloglu, I. Bahar and B. Erman, Gaussian dynamics of folded proteins. *Phys. Rev. Lett.,* **79**, 3090, (1997).

[23] K. Hinsen, Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33**, 417, (1998).

[24] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J.,* **80**, 505, (2001).

[25] F. Tama and Y. H. Sanejouand, Conformational change of proteins arising from normal mode calculations. *Protein. Eng.,* **14**, 1 (2001).

[26] O. Kurkcuoglu, R. L. Jernigan and P. Doruker, Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymers*, **45**, 649, (2004).

[27] P. Doruker, R. L. Jernigan and I. Bahar, Dynamics of large proteins through hierarchical levels of coarse grained structures. *J. Comput. Chem.,* **23**, 119, (2002).

[28] S. Miyazawa and R. L. Jernigan, Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, **18**, 534, (1985).

[29] I. Bahar and Jernigan, R. L., Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.*, **266**, 195, (1997).

[30] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Brooks Cole, (1976).

[31] S. Alexander and R. Orbach, Density of states of fractals: 'fractons'. *J. Phys. (Paris) Lett.,* **43**, L625, (1982).

[32] R. Rammal and G. Toulouse, Random walks on fractal structures and percolation clusters. *J. Phys. (Paris) Lett.,* **44**, L13, (1983).

[33] R. Burioni, D. Cassi, F. Cecconi and A. Vulpiani, Topological thermal instability and length of proteins. *Proteins* **55**, 529, (2004).

[34] L. Yang, E. Eyal, C. Chennubhotla, J. Jee, A. Gronenborn and I. Bahar, Insights into Equilibrium Dynamics of Proteins from Comparison of NMR and X-ray Data with Computational Predictions. *Structure,* **15**, 741, (2007).

[35] R. Peierls, *Helv. Phys. Acta*, **2**, 81, (1934).

[36] P. G. De Gennes, *Scaling Concepts in Polymer Physics*, Cornell Univ. Press, (1979).

[37] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, (1988).

[38] L. S. Schulman, *Techniques and Applications of Path Integration,* John Wiley & Sons, (1981).

[39] M. Kardar, *Statistical Physics of Fields*, Cambridge University Press, (2007).

[40] P.G. Doyle and J.L Snell, *Random walks and electric networks*, Carus Mathematical Monographs, (1984).

[41] T. Nakayama and K. Yakubo, *Fractal Concepts in Condensed Matter Physics,* Springer, (2003).

[42] R. Durrett, *Essentials of Stochastic Processes*, Springer-Verlag, (1999).

[43] D. Joseph, G. A. Petsko, and M. Karplus, Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science,* **249**, 1425, (1990).

[44] P. J. Steinbach *et al.*, Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry,* **30**, 3988, (1991).

[45] B. F. Rasmussen, A. M. Stock, D. Ringe, and G. A. Petsko, Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature,* **357**, 423, (1992).

[46] A. Henzler-Wildman *et al.*, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature,* **450**, 913, (2007).

[47] S. Reuveni, R. Granek and J. Klafter, Proteins: coexistence of stability and flexibility. *Phys. Rev. Lett.* **100**, 208101, (2008).

[48] M. de Leeuw, S. Reuveni, J. Klafter and R. Granek, Coexistence of flexibility and stability of proteins: an equation of state. *PLoS ONE,* **4**(10), (2009).

[49] S. Reuveni, R. Granek and J. Klafter, Anomalies in the vibrational dynamics of proteins are a consequence of fractal like structure. *Proc. Natl. Acad. Sci.,* **107**, 13696, (2010).

[50] H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinson and S. R. Kurtz, Fractal form of proteins. *Phys. Rev. Lett.,* **45**, 1456, (1980).

[51] S. G. Lushnikov, A. V. Svanidze and I. L. Sashin, Vibrational density of states of hen egg white lysozyme. *JETP Letters,* **82**, 30, (2005).

[52] X. Yu and D. M. Leitner, Anomalous diffusion of vibrational energy in proteins. *J. Chem. Phys.,* **119**, 12673, (2003).

[53] D. M. Leitner, Energy Flow in Proteins. *Annu. Rev. Phys. Chem.,* **59**, 233, (2008).

[54] G. Haran, E. Haas, B. K. Szpikowska and M. T. Mas, Domain motions in phosphoglycerate kinase: determination of interdomain distance distributions by site-specific labeling and time-resolved fluorescence energy transfer. *Proc. Natl. Acad. Sci.,* **89**, 11764, (1992).

[55] K. Arora and C. L. Brooks III, Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci.,* **104**, 18496, (2007).

[56] I. Bahar, C. Chennubhotla and D. Tobi, Intrinsic Enzyme Motions in the Unbound State and Relation to Allosteric Regulation. *Curr. Opin. Struct. Biol.*, **17(6)** 633, (2007).

[57] R. Granek and J. Klafter, Fractons in Proteins: Can They Lead to Anomalously Decaying Time Autocorrelations? *Phys. Rev. Lett.,* **95**, 098106, (2005).

[58] C. Chennubhotla and I. Bahar, Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. *PLOS Comput. Biol.,* **3**, e172, (2007).

[59] T. Nakayama *et al.*, Dynamical properties of fractal networks: Scaling, numerical simulations, and physical realizations. *Rev. Mod. Phys.,* **66**, 381, (1994).

[60] M. E. Cates, Brownian dynamics of self-similar macromolecules. *J. Phys. France,* **46**, 1059, (1985).

[61] J. E. Martin *et al.*, Viscoelasticity near the sol-gel transition. *Phys. Rev. A,* **39**, 1325, (1989).

[62] A. G. Zilman and R. Granek, Dynamics of fractal sol-gel polymeric clusters, *Phys. Rev. E,* **58**, R2725, (1998).

[63] C. M. Sorensen, Light scattering by fractal aggregates: a review. *Aerosol Science and Technology* **35**, 648, (2001).

[64] E. L. Aiden *et al.*, Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science **326**, 289, (2009).

[65] A. Y. Grosberg, S.K. Nechaev and E.I. Shakhnovich, The role of topological constraints in the kinetics of collapse of macromolecules. *J. Physique* (France), **49**, 2095, (1988).

[66] A. Y. Grosberg, Y. Rabin, S. Havlin and A.Neer. Crumpled Globule Model of the Three-Dimensional Structure of DNA. *Europhys. Lett.,* **23**, 373, (1993).

[67] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*, Cambridge Uni. Press, (1995).

[68] S. Reuveni, J. Klafter and R. Granek, Dynamic Structure Factor of Vibrating Fractals. *Phys. Rev. Lett.,* **108**, 068101, (2012).

[69] M. Monkenbusch and D. Richter, High resolution neutron spectroscopy - a tool for the investigation of dynamics of polymers and soft matter. *C. R. Physique,* **8**, 845, (2007).

[70] M. Monkenbusch, D. Richter and R. Biehl, Observation of protein domain motions by neutron spectroscopy. *ChemPhysChem,* **8**, 1188, (2010).

[71] J. Lal, P. Fouquet, M. Maccarini, and L.Makowski, Neutron Spin-Echo Studies of Hemoglobin and Myoglobin: Multiscale Internal Dynamics. *J. Mol. Biol.,* **397**, 423, (2010).

[72] S. Reuveni, J. Klafter and R. Granek, Dynamic Structure Factor of Vibrating Fractals: Proteins as a Case Study. *Phys. Rev. E.,* **85**, 011906, (2012).

[73] S. Havlin and D. ben-Avraham, Diffusion in disordered media. *Adv. Phys.,* **36**, 695, (1987)

[74] D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems,* Cambridge Univ. Press, (2000).

[75] H. C. Tuckwell, *Introduction to Theoretical Neurobiology*, Cambridge Univ. Press, (1988).

[76] A. L. Lloyd and R. M. May, Epidemiology — how viruses spread among computers and people. *Science,* **292**, 1316, (2001)

[77] O. Bénichou, M. Coppey, M. Moreau, P. H. Suet and R. Voituriez, Optimal search strategies for hidden targets. *Phys. Rev. Lett.,* **94**, 198101, (2005).

[78] O. Bénichou, C. Loverdo, M. Moreau and R. Voituriez, Two-dimensional intermittent search processes: An alternative to Lévy flight strategies. *Phys. Rev. E,* **74**, 020102, (2006).

[79] M. F. Shlesinger, Mathematical physics: Search research. *Nature,* **443**, 281, (2006).

[80] I. Eliazar, T. Koren and J. Klafter, Searching circular DNA strands. *J. Phys. Condens. Matter.,* **19**, 065140, (2007).

[81] S. Condamin, O. Benichou, V. Tejedor, R. Voituriez and J. Klafter, First-passage times in complex scale-invariant media. *Nature,* **450**, 77, (2007).

[82] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer and R. Voituriez, Geometry-controlled kinetics. *Nature Chemistry,* **2**, 472, (2010).

[83] S. Reuveni, R. Granek and J. Klafter, Vibrational shortcut to the mean-first-passage-time problem. *Phys. Rev. E.,* **81**, 040103(R), (2010).

[84] S. Reuveni, R. Granek and J. Klafter, General Mapping Between Random Walk and Thermal Vibrations in Elastic Networks: Fractal Networks as a Case Study. *Phys. Rev. E.,* **82**, 041132, (2010).

[85] A. Banerji and I. Ghosh, Fractal symmetry of protein interior: what have we learned? *Cell. Mol. Life Sci.,* **68**, 2711, (2011).

[86] S.C. Kou and X.S. Xie, Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule. *Phys. Rev. Lett.,* **93**, 180603, (2004).

[87] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Observation of a Power-Law Memory Kernel for Fluctuations within a Single Protein Molecule. *Phys. Rev. Lett,* **94**, 198302, (2005).

[88] P. Senet, G. G. Maisuradze, C. Foulie, P. Delarue, and H. A. Scheraga, How main-chains of proteins explore the free-energy landscape in native states. *Proc. Natl. Acad. Sci.,* **105**, 19708, (2008).

[89] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, Subdiffusion in Peptides Originates from the Fractal-Like Structure of Configuration Space. *Phys. Rev. Lett.,* **100**, 188103, (2008).

[90] E. Haas, The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. *ChemPhysChem*, **6**, 858, (2005).

[91] J. H. Roh *et al.*, Influence of Hydration on the Dynamics of Lysozyme. *Biophys. J.,* **91**, 2573, (2006).

[92] S. Khodadadi *et al.*, The origin of the dynamic transition in proteins. *J. Chem. Phys.,* **128**, 195106, (2008).

[93] L. Hong, N. Smolin, B. Linder, A. P. Sokolov and J.C. Smith, Three Classes of Motion in the Dynamic Neutron-Scattering Susceptibility of a Globular Protein. *Phys. Rev. Lett.,* **107**, 148102, (2011).

[94] G. R. Kneller and K. Hinsen, Fractional Brownian dynamics in proteins. *J .Chem. Phys.*, **121**, 10278, (2004).

[95] V. Calandrini, V. Hamon, K. Hinsen, P. Calligari, M. C. Bellissent-Funel and G.R. Kneller, Relaxation dynamics of lysozyme in solution under pressure: Combining molecular dynamics simulations and quasielastic neutron scattering. *Chem. Phys.*, **345**, 289, (2008).

[96] A. Blumen, Ch. von Ferber, A. Jurjiu, and Th. Koslowski, Generalized Vicsek fractals: Regular hyperbranched polymers. *Macromolecules,* **37**, 638, (2004).

[97] A. Blumen, A. Jurjiu, Th. Koslowski, and Ch. von Ferber, Dynamics of Vicsek fractals, models for hyperbranched polymers. *Phys. Rev. E.,* **67**, 061103, (2003).

[98] R. Granek, Proteins as fractals: Role of the hydrodynamic interaction. *Phys. Rev. E,* **83**, 020902(R), (2011).

[99] J. Mattsson *et al.*, Soft colloids make strong glasses. *Nature,* **462**, 83, (2009).

[100] A.H. Krall and D.A. Weitz, Internal Dynamics and Elasticity of Fractal Colloidal Gels. *Phys. Rev. Lett.,* **80**, 778, (1998).

[101] J. G. McNally and D. Mazza, Fractal geometry in the nucleus. *The EMBO Journal,* **29**, 2, (2009).

[102] A. J. Einstein, H. S. Wu, M. Sanchez, J. Gil., Fractal characterization of chromatin appearance for diagnosis in breast cytology. *J. Pathol.,* **185**, 366, (1998).

[103] V. Bedin, R. L. Adam, C. S. Bianca, L. Gilles and K. Metze, Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma. *BMC Cancer,* **10**, 260, (2010).

[104] I. Bronstein, Y. Israel, E. Kepten, S. Mai, Y. Shav-Tal, E. Barkai and Y. Garini, Transient Anomalous Diffusion of Telomeres in the Nucleus of Mammalian Cells. *Phys. Rev. Lett.,* **103**, 018102, (2009).