Warszawa, May 23rd, 2016

Dr Joanna Ida Sułkowska

Interdisciplinary Laboratory of Biological Systems Modelling
Centre of New Technologies

Chair for Theoretical Chemistry and Crystallography
Faculty of Chemistry

University of Warsaw

AUTOREFERAT

Contents

1. Name

Joanna Ida Sułkowska

2. Diplomas and scientific degrees

Scientific degree: PhD in Physical Sciences
Institution: Institute of Physics, Polish Academy of Sciences
Year: 2007 (with distinction)
Title of PhD thesis: “Stretching and folding proteins in coarse grained models”
Advisor: Prof. Marek Cieplak

MSc. diploma
Institution: Faculty of Physics, University of Warsaw
Year: 2003
Advisors:

- Prof. Christoph F. Schmidt (Vrije Universiteit Amsterdam)
- Prof. Marek Cieplak (Institute of Physics, Polish Academy of Sciences)

3. Information about employment in scientific institutions

1. **University of Warsaw**, Centre of New Technologies
A group leader – a head of the "Interdisciplinary Laboratory of Biological Systems Modelling", from 2014

2. **University of Warsaw**, Faculty of Chemistry
Adjunct (assistant professor), from 2013
Scientific and technical researcher, 2012-2013

3. **Rice University, Houston,** Faculty of Chemistry
Visiting faculty, 2012-2013

4. **University of California San Diego,** Center for Theoretical Biological Physics
Postdoctoral researcher, 2008-2012

5. **University of San Diego,** Faculty of Biochemistry
Visiting faculty, 2012-2013

6. **Polish Academy of Sciences**, Institute of Physics
Adjunct (assistant professor), 2008-2010

4. Scientific achievement, in the sense of article 16, paragraph 2 of the *Act on academic degrees and academic title and degrees and title in art* (Dz. U. nr 65, poz. 595 ze zm.)

A) Title of the scientific achievement

Classification and the free energy landscape of entangled proteins: knots, slipknots and lassos

B) Publications included in the scientific achievement

H1. P. Dąbrowski-Tumanski, W. Niemyska, P. Pasznik, J.I. Sułkowska,
"LassoProt: server to analyze biopolymers with lassos",
Nucleic Acids Res. (2016), doi: 10.1093/nar/gkw308.
Internet database and webserver: ***http://LassoProt.cent.uw.edu.pl***

H2. P. Dąbrowski-Tumanski, A.I. Jarmolińska, J.I. Sułkowska,
"Prediction of the optimal set of contacts to fold the smallest knotted protein",
J. Phys. Cond. Mat. (2015) 27(35):354109, doi: 10.1088/0953-8984/27/35/354109.

H3. M. Jamroz, W. Niemyska, E.J. Rawdon, A. Stasiak, K.C. Millett, P. Sułkowski, J.I. Sulkowska,
"KnotProt: a database of proteins with knots and slipknots",
Nucleic Acids Res. (2014), 43: D306-D314.
Internet database and webserver: ***http://KnotProt.cent.uw.edu.pl***

H4. E. Haglund, J.I. Sulkowska, J.K. Noel, H. Lammert, J.N. Onuchic, P.A. Jennings,
"Pierced Lasso Bundles are a New Class of Knot Motifs",
PloS Comput. Biology (2014) 19,10(6):e1003613.

H5. P. Dabrowski-Tumanski, S. Niewieczerzal, J.I. Sulkowska,
"Determining Critical Amino Acid contacts for knotted protein",
TASK Quarterly (2014) 18 No 3, 323–337.

H6. E.J. Rawdon, K.C. Millett, J.I. Sułkowska, A. Stasiak,
"Knot localization in proteins",
Biochemical Society Transactions (2013) 41(2):538-41. doi: 10.1042/BST20120329.

H7. K.C. Millett, E.J. Rawdon, A. Stasiak, J.I. Sułkowska,
"*Identifying knots in proteins*",
Biochemical Society Transactions (2013) 41(2):533-7, doi: 10.1042/BST20120339.

H8. T. Andrews, D.T. Capraro, J.I. Sułkowska, J.N. Onuchic, P.A. Jennings,
"Hysteresis as a Marker for Complex, Overlapping Landscapes in GFP and Other Proteins",
J. Phys. Chemistry Letters (2012) Jan 3;4(1):180-188.

H9. J.K. Noel, J.N. Onuchic, J.I. Sulkowska,
***"Knotting in explicit solvent"*,**
J. Phys. Chem. Letters (2013) 4(21), 3570-3573.

H10. J.I. Sułkowska, J.K. Noel, C. A. Ramírez-Sarmiento, E.J. Rawdon, K.C. Millett, J. N. Onuchic,
"*Knotting pathways in proteins*",
Biochemical Society Transactions (2013) 41(2):523-7, doi: 10.1042/BST20120342.

H11. J.I. Sułkowska*, J. K Noel*, J. N. Onuchic,
"*Energy landscape of knotted protein folding*",
Proc. Natl. Acad. Sci. USA (2012) doi: 10.1073/pnas.1201804109.

H12. J.I. Sułkowska*, E. J. Rawdon*, K. C. Millett, J. N. Onuchic, A. Stasiak,
"*Conservation of complex knotting and slipknotting patterns in proteins*",
Proc. Natl. Acad. Sci (USA), (2012) 109(26): E1715-23.

H13. E. Haglund, J.I. Sułkowska, Zhao He, Gen-Sheng Feng, P. Jennings, J. N. Onuchic,
"*The unique cysteine knot regulates the pleotropic hormone leptin*",
PloS One (2012) 7(9) e45654.

H14. J.K. Noel, J.I. Sułkowska, J.N. Onuchic,
"Slipknotting upon native-like loop formation in a trefoil knot protein",
Proc. Natl. Acad. Sci. USA (2010) 107, 15403.

H15. J.I. Sułkowska, P Sułkowski, P Szymczak, M Cieplak,
"Untying knots in proteins",
J. Am. Chem. Soc. 132 (40) (2010) 13954-13956.

H16. D. Bölinger, J.I. Sułkowska, H.P. Hsu, L.A. Mirny, M. Kardar, J.N. Onuchic, P. Virnau,
"A Stevedore's protein knot",
PLoS Comput Biol. (2010) 6, e1000731.

H17. J.I. Sułkowska, P. Sułkowski, J.N. Onuchic,
"Dodging the crisis in protein folding with knots",
Proc. Natl. Acad. Sci. USA (2009) 106, 3119-3124.

H18. J.I. Sułkowska, P. Sułkowski, J.N. Onuchic,
"Jamming proteins with slipknots and their free energy landscape",
Phys. Rev. Lett. (2009) 103 268103.

H19. J I. Sułkowska, P Sułkowski, P Szymczak, M Cieplak,
"Stabilizing effect of knots on proteins - How knots influence properties of proteins",
Proc. Natl. Acad. Sci. USA 105 (2008) 19714-19719.

C) Description of the scientific aim of the above publications and the results

Table of Contents

B Detailed analysis of folding of the smallest knotted protein [H2,H5,H9-10,H14]
C Energy landscape of proteins with lassos [H4,H13]

V The free energy landscape – unfolding and untying of knots and slipknots

A Knotted proteins in denatured state and a model to untie them [H8,H15,H19]
B Unique metastable properties of proteins with slipknots [H18]

I Introduction

Proteins are fundamental ingredients of living organisms. In order to understand molecular foundations of their function one needs to uncover their structure and dynamics, as well as links between structure and dynamics. Analysis of the free energy landscape is a very important method of achieving these goals. The structure of around 91% of known proteins is uniquely determined by their primary, secondary, and tertiary form, and their free energy landscape can be, quite accurately, determined experimentally and theoretically (e.g. in computer simulations, by means of coarse grained models, as well as all atom models for smaller biomolecules). However in last years it has been revealed – in particular in the course of studies conducted by myself – that the remaining 9% of proteins possess nontrivial topology (i.e. are entangled): 1.5% of those proteins possess knots and slipknots [H3], and 7.5% possess lassos [H1] (examples of knots, slipknots and lassos are shown in figure 1). A standard description of spatial configurations (in terms of primary, secondary, and tertiary structures) of such entangled proteins turns out to be insufficient, and various techniques and theories that have been used to date need to be modified in a way which would take into account the presence of entanglement.

In this work I present foundations of a new research field, devoted to studies of entangled proteins. Among others I discuss properties of proteins with knots, slipknots and lassos identified in my research, and characterize their free energy landscape, using tools developed also by myself. This work consists of a series of publications devoted to: formulation of various methods and their application to the classification of entangled proteins; development of models and tools, both theoretical and experimental, to determine physical and chemical properties that shape free energy landscape of such proteins; understanding how entanglement in proteins affects their evolution and biological function. These results provide foundations of a new discipline, devoted to research on entangled proteins, which currently is being actively developed.

Figure 1: *Schematic representation of various entangled structures discussed in this work (left to right): knots, slipknots, and lassos.*

Before presenting the results of my research, I will first briefly summarize entangled structures in proteins – knots, slipknots, and lassos, shown schematically in figure 1 – which I have identified in my work.

According to a mathematical definition, a **knot** is a closed, non-self-intersecting curve embedded in a three-dimensional space. A branch of mathematics known as knot theory provides tools to classify and describe properties of knots. Examples of those tools are so called knot invariants, i.e. some relatively simple mathematical objects (e.g. numbers, polynomials, etc.), which can be assigned algorithmically to a given knot, and then used to distinguish or classify knots. A famous example of such an invariant is the Jones polynomial – for its discovery Vaughan Jones was awarded with the Fields medal.

Knots play an important role not only in mathematics, but also in other disciplines (e.g. physics and biology), and their existence may have various consequences (e.g. effectively packed DNA in a capsid should be knotted). However until the last decade of the 20^{th} century it was believed that protein chains cannot be knotted, in particular because achieving such configurations would be very unfavorable energetically. Such a statement was formulated e.g. in an important work by Mansfield, published in Nat. Struct. Biol. in 1994, where the first review of experimentally determined protein structures was conducted (1). In this review only one knotted protein was identified, which contained a shallow knot, i.e. a knot with only a few amino acids (represented by black curves in figure 1) extending beyond its core (represented by a blue curve). Such shallow knots can be spontaneously untied by thermal fluctuations, so their existence was regarded as unimportant. Moreover analysis of knotting in proteins leads to mathematical problems: knots can be rigorously defined only on closed curves (without free ends), while protein chains are open curves (and have free ends). For this reasons defining knots in proteins is a subtle task, and new theoretical tools are necessary to conduct their analysis and classification. Such tools, previously unknown, have been developed by myself and my collaborators, as described in the rest of this work. As will be shown in what follows, knots in open curves are very intriguing objects, and their role in biology may have fundamental meaning for many life processes.

The first protein with a deep knot, i.e. a knot with at least 15 amino acids (represented by black curves in figure 1) extending beyond its core, was identified (by means of a simple theoretical KMT algorithm) by Taylor (2) in the year 2000. In the years 2006-2007 Grosberg and Virnau (3,4) showed, that when free ends of a knot in a native conformation are sufficiently far from its core, then they can be connected in a unique way and a type of a knot can be determined (by the same methods as for closed curves). Subsequently other knotted structures were identified, and the paradigm in thinking about protein complexity changed. Furthermore, while many new knotted proteins were identified, it turned out that statistically they are formed much less frequently than in other (bio)polymers. In this work I will explain these observations by invoking arguments based on evolutional selection, specific properties of knots on open chains, and problems with experimental analysis of entangled proteins.

The second class of entangled structures analyzed in this work are **slipknots**. These configurations consists of two loops, with one loop partly threaded through the second loop [H17-H18], as shown schematically in figure 1 (in the middle). Cutting several amino acids from one free end transforms a slipknot into an open knot. After connecting two ends of a slipknot (without introducing any new crossings), its topology is trivial from the knot theory viewpoint. Nonetheless, slipknots in proteins turn out to possess very intriguing mechanical properties [H18]. Even though formation of slipknots may seem more likely and energetically more favorable than formation of knots, first six proteins with slipknot were discovered only in the year 2006 (5). There are many interesting questions concerning slipknots: how often are they formed in proteins, how to determine and classify them, how complex can they be, how their presence affects mechanical and thermodynamical properties of proteins and their biological function. I will answer these questions in the following chapters of this work.

Proteins with **lassos** constitute the third class of entangled structures discovered and analyzed in this work. According to the definition formulated in [H1], lasso is a structure that contains a covalent loop (closed by a cysteine, amide, or other bridge), crossed by at least one free end of the

structure. The simplest example of a lasso is shown in figure 1 (right). Identification and classification of lassos was motivated by:
1) properties of proteins with cysteine bridges, which can form so called cysteine knots; analysis of these proteins revealed, that their complicated geometry improves their stability,
2) existence of complex chemical compounds, whose local connections, e.g. involving metal ions, may form nontrivial topologies,
3) mini-proteins, which consist of a small closed loop (made of around 12 amino acids), through which one terminus is threaded; such a conformation behaves like a "plug", which can block biological activities of bacterial RNA polymerase,
4) analysis of literature and research results, showing that closed polymer chains may form various types of entanglement, such as circular super-coiled DNA structures,
5) the role of the theory of knots and links in understanding of biological functions of topoisomerazes and recombinazes, i.e. two classes of enzyms, which control the topology of DNA and RNA; pioneering results (in 1980) of two mathematicians Ernst and Summers on the "tangle calculus", as well as by M. Vazquez who used the theory of links to describe entanglement in replicated bacterial chromosoms, were crucial for the development in this area.

The existence of entangled structures – knots, slipknots, and lassos – described above shows, that spatial conformation of a protein backbone chain may be much more complex than it had been believed before, and in particular it may affect biological functions of proteins. Analysis of the structure and the role of entanglement in proteins requires a new, interdisciplinary approach, which is developed and presented in this work.

The existence of entangled structures is also important from the viewpoint of methods for protein structure prediction, as well as evolutional processes. According to the basic assumption of structural biology, the sequence of amino acids determines the spatial structure of a protein (7). Analysis of sequences of amino acids, supported with various bioinformatics, physical, statistical, and chemical methods, enables to predict protein structure; in particular such methods are used in the CASP competition (The Critical Assessment of Protein Structure Prediction). The discovery of entangled proteins raises various questions concerning methods to predict protein structure – when predicting this structure, it is desirable to verify whether the resulting topology is correct. In view of the existence of entangled structures, also the following question becomes important: to what extent is the entanglement (the presence of knots, slipknots, or lassos) preserved evolutionally? First works by Bujnicki and Micheletti (8,9) suggest, that a protein with a knot evolved from an unknotted protein. However these works do not explain whether knots that we know to date survived evolutional selection e.g. due to their crucial role for biological function. The impact of topology on the evolution of proteins – and vice versa – is the next topic considered in this work.

The free energy landscape plays a crucial role in the analysis of entangled proteins. In particular this landscape encodes information about relations between proteins' structure and dynamics. The energy landscape describes both local changes in the protein structure, related to the activation energy necessary for enzymes to initiate chemical reactions, as well as global conformational changes – such as folding, degradation, mechanical unfolding – whose character is determined by physical and chemical properties of amino acids and the geometry of the system.

In general the process of protein folding is different than a multi-step transformation in organic synthesis or metabolic processes – it is not determined by a specific sequence of steps, and it is not characterized by a large number of stable intermediate states (11). Protein folding may be described by several main routes, which contain intermediate states only in case of more complex proteins. In the analysis of protein folding by means of coarse grained models, several main approaches are based either on proteins' native states (12,13,14), or statistical potentials (14,15,16). In this work I will focus on the approach based on the native structure of proteins, which has been successful in many theoretical and experimental developments. According to this approach, in most cases the

energy landscape takes form of a funnel, while proteins can be treated as minimally frustrated heteropolymers, which attain native conformations after following one of their folding routes, using only attracting native contacts (11). The folding process resembles diffusion in a relatively smooth energy landscape, where a set of possible configurations shrinks upon approaching the native configuration. Folding of such heteropolymers is quite successfully described by coarse grained Go-like models, also known as Structure Based Models (SBM), which use the information of the native structure. It is worth noting, that formulation of multiscale theoretical models to describe proteins and other complex chemical structures was recognized by the Nobel prize in chemistry in the year 2013.

The assumption of the minimal frustration is currently widely used to simulate folding processes, modeling, and predicting protein structure. Some doubts regarding this postulate – based e.g. on the Levinthal paradox (17) – have been raised in past two decades. However results obtained with new supercomputers, e.g. the Anton (constructed by D. Shaw), provide new convincing arguments that the above postulate is true. Simulations in an explicit solvent showed, that the same contacts are used to fold proteins both in SBM models, and in models with explicit solvent (18). Moreover for most proteins the number of native contacts provides a good reaction coordinate to describe folding processes and intermediate states (19). The hypothesis that proteins can be regarded as minimally frustrated heteropolymers formed in the course of evolution, has been recently strongly supported by the analysis of coevolution of pairs of amino acids, within the Direct Coupling Analysis (DCA) method (11). The coevolution of pairs of amino acids shows, that even though minimally frustrated sequences are very rare, they are distributed with sufficient density in the space of sequences, which enables identification of a proper minimally frustrated sequence under sufficient evolutional pressure (11). One should however note, that for some proteins such an energy landscape is not so ideal (19).

To sum up, the approach based on the native structure of proteins and the funnel landscape theory is efficient and useful from several perspectives. However an important question arises, whether the funnel landscape theory and the hypothesis of minimal frustration is relevant also for proteins with nontrivial topology, whose configuration space is restricted by topological constraints. In this work I will provide an answer to this question, formulate a theory of folding and unfolding of proteins with nontrivial topology, and show its efficiency.

Independently of theoretical methods, experimental analysis of proteins is of course also very important. In the last decade we have witnessed an unprecedented development in experimental techniques, and currently we can use optical tweezers or Atomic Force Microscopy to analyze even single molecules (20,21). Moreover application of computer methods of molecular dynamics in all atom models (22) and coarse grained models (23), together with a theory by Kramers, Bell, Shabo, and Dudko, enables interpretation of experimental data and their modeling in a wide range of the energy landscape, including details that are often unavailable in the thermodynamical approach. Mechanical unfolding of single molecules is currently a precious source of knowledge about the nature of DNA, RNA, and proteins.

Applications – also in my works – of the methods of mechanical unfolding shows, that the geometry of a system is the main factor that determines the process of mechanical denaturation (23). However also in this case it is not obvious whether previously known theory and methods can describe properly mechanical unfolding of proteins, e.g. the process of their degradation. Other related and important new questions arising in this context are as follows: how to untie the entangled structures, what is the role of knots, do knots, slipknots and lassos introduce additional stability. The answers to these questions will be provided in the rest of this work.

The above discussion clearly shows, that new methods and techniques need to be developed to analyze 9% of known protein structures. Achieving this goal requires an interdisciplinary approach.

It is also quite likely, that more and more entangled proteins will be discovered in the future, and in consequence their role may turn out to be even more important.

In this work an interdisciplinary approach involving methods from physics, biophysics, biochemistry, and knot theory is presented, with the aim of explaining the role of nontrivial topology for functions of proteins. The methods developed in this work form a basis of a new emerging research area, devoted to entangled proteins. The main aims and results of this work are as follows:

- construction of tools that enable theoretical description of entanglement in proteins, a discovery of new proteins with contain knots and lassos, and conducting the first ever classification of proteins with knots [H6,H7,H12], slipknots [H17,H18] and lassos [H1];
- revealing the evolutional and biological meaning of entangled proteins [H1, H4,H11,H12,H16,H19];
- understanding the process of self-tying of proteins with deep knots and slipknots [H11,H16,H17];
- reconstruction of the energy landscape of the smallest knotted protein by means of simulations in all atom and coarse grained models [H2,H5,H9,H10,H14];
- understanding of the process of self-tying of proteins with lassos [H4, H13] and characterization of differences in untying of proteins of various topology and the same fold [H19];
- proving the existence of knots in denatured states of proteins [H8], derivation of a formula for a probability of knot untying [H15], and identification and analysis of unique metastable properties of proteins with slipknots [H18].

The plan of this work is as follows. In chapter II the methods of analysis of nontrivial topology in proteins are presented, together with structures identified in the course of my work. In chapter III I will show that topology of knots and slipknots is evolutionally preserved, and I will present the meaning and the role of entanglement for biological function and mechanical properties of knots. In chapter IV I will characterize the mechanism of formation of knots in proteins. In chapter V I will prove that knots in denatured states of proteins exist, I will explain how they can be untied, and characterize unique behavior of proteins with slipknots upon mechanical untying.

II Identification of entangled structures in proteins: knots, slipknots, lassos, and their classification

In this section I will present entangled structures and models for their identification [H1,H6,H7,H12,H18]. I will also provide a short introduction to knot theory and a summary of knotted proteins discovered earlier (before my work). Properties of proteins with knots and slipknots are summarized in sections **A** and **B**. While some knotted proteins had indeed been found before my work, it should be stressed that analysis of proteins with lassos (summarized in sections **C** and **D**) is a completely new topic.

II.A Matrix model description of knots and slipknots

To start with we recall, that knots are defined as closed, non-self-interacting curves embedded in three-dimensional space. Knots may be classified using so called knots invariants – some other mathematical objects (e.g. numbers or polynomials) which depend only on the topology of a given knot, and not on details of its shape. If certain invariants for two knots are different, it means that these knots are topologically inequivalent, and it is not possible to transform smoothly one onto the other. The simplest knot invariant is defined as a minimal number of crossings in a projection of a

given knot on a plane. This invariant is used in the standard notation of knot types – more precisely, it is combined with another number (written in subscript), which labels consecutive knots with the same number of crossings. The simplest knot is called an unknot or a trivial knot, denoted 0_1 – this is just an unknotted loop and its number of crossings is zero. The first nontrivial knot is called a trefoil; it is characterized by three crossings and denoted 3_1; other more complicated knots include the figure-eight knot with four crossings, denoted 4_1, two knots with five crossings denoted 5_1 and 5_2, knots 6_1, 6_2 and 6_3 with six crossings; the numbers of knots with larger numbers of crossings are rapidly growing. More complicated invariants, which enable to distinguish more knots, take form e.g. of polynomials in one or more variables, such as the Alexander polynomial, Jones polynomial or HOMFLY polynomial. In this work we use such polynomial knot invariants to identify knots in proteins.

While knot theory provides methods to characterize topological configurations on closed curves, its application to proteins, which form open chains, requires certain generalization. To identify a knot in a protein one needs first to close it (connect its ends in order to form a closed curve), and details of this closing procedure – which is not unique – may affect the resulting knot type and its interpretation for global and local mechanical, thermal, and biological properties. Moreover a detailed description of the knot's structure requires considering how all subchains of the backbone chains are knotted. All these arguments are inspired by earlier works, in particular by Grosberg, Taylor, Yeates, and my own considerations.

In [H12] I developed a method to determine statistical entanglement (i.e. the most probable knot type) for each subchain of a given peptide chain. By a subchain I understand a peptide chain with a number of amino acids cut from the N-terminal side or C-terminal side. This method is called a matrix model, because the most probable knot type of each chain can be presented in a form of a matrix, or a diagram representing it, which is also called a "knot fingerprint". Each element of such a matrix, corresponding to a given subchain, is determined independently. This model solves the problem of non-uniqueness of determination of knots on open curves, and moreover enables identification and characterization of slipknots. Knot fingerprint provides a visual identification of various subchains, which represent various knot types. Details of the construction of knot fingerprints, the chain closure methods, and other advantages of this matrix model are summarized in two papers [H6,H7].

Furthermore, the matrix model enables to determine a landscape of topological changes in proteins, e.g. during folding processes. An optimized version of a matrix model that enables to determine topological landscape and knot fingerprints is currently widely used as an online server *KnotProt* (***http://KnotProt.cent.uw.edu.pl***) [H3].

II.B Classification of knots and slipknots

One of important results of my work is the classification of all proteins with knots and slipknots. Using the matrix model described above to the analysis all proteins deposited in PDB, I characterized types and arrangements of possible entanglements [H9]. For example I showed that subchains of a protein with 6_1 knot, which I discovered in [H13], form also another 6_1 knot, as well as 4_1 knot and 3_1 knot. These knots can be detected upon removing amino acids from the termini of the peptide chain. Another example is a slipknot that includes 3_1,4_1,3_1 knots, which are also detected by removing amino acids either from C-terminus (this leads to 3_1 knot), N-terminus (this leads to 4_1 knot), or both termini simultaneously (this gives another 3_1 knot). In [H9] I also showed (thereby confirming earlier predictions [31,32]), that a topological change is possible only via a slipknot configuration, which transforms a family of knots known as twist knots into the unknot, upon a reduction of only one crossing.

Characterization of entanglement in proteins by means of the matrix model is important for several reasons: first, it shows that each protein has its own, unique knot fingerprint; second, it

enables to clearly distinguish knots from slipknots (some examples will be given below); third, it enables to classify entanglements in proteins – i.e. proteins with similar fingerprints can be classified as belonging to the same topological class. Fourth, it enables to identify how a given entangled structure arises in various proteins, thereby suggesting evolutional role of knots, which will be summarized in section III. A list of knotted proteins which I identified and detailed description of their motives are described in [H9].

Another important result of this work is a completion of a topological classification of proteins [H9] and construction of an online database *KnotProt* – the first database ever which contains full information about types of knotting of all proteins [H2]. This database is available at the address ***http://KnotProt.cent.uw.edu.pl***. Details of this classification and characterization of knotted proteins are presented in [H9] and in reviews [H5,H6]. In those publications I identified 16 knotting motifs in proteins (Table 1, [H9]), which are denoted by a letter K or S (denoting whether the full chain forms a knot or a slipknot), followed by a list of all knots identified in various subchains. The largest two groups are formed by proteins with a single 3_1 knot, or a single slipknot also of type 3_1. The third largest class consists of proteins with a slipknot of type $S3_14_13_1$, which has been identified in membrane proteins. Entanglement in membrane proteins is particularly interesting – using *KnotProt* I identified 3_1 knot in four membrane proteins (knots in membrane proteins had not been identified before!), and found that as much as around 20% of those proteins possess slipknots. *KnotProt* database is a source of knowledge about nontrivial entangled motifs in proteins, and includes their structural and biological classification. *KnotProt* database is automatically updated every week (taking into account newly deposited proteins in PDB), and therefore it provides the most up to date source of knowledge about proteins with knots and slipknots.

It should be stressed, that *KnotProt* can also be used as a server for analysis of open polymers (determined e.g. from molecular dynamics simulations). This server has many applications, for example it can be used to verify whether a topology of experimentally identified structures is correct (e.g. in Cryo-electron microscopy, cryo-EM, which often produces incorrect entanglement), or to support analysis conducted in the CASP competition. *KnotProt* is also very often used to analyze entanglement in proteins and polymers obtained in computer simulations – to my best knowledge it is the only commonly used tool to analyze dynamics of entangled biomolecules.

II.C Lassos and minimal surfaces

Another result of this work is a formulation of a method to identify lasso structures. In [H1] I defined lasso configurations and formulated a model of *minimal surfaces* for their identification. Lasso configurations are formed in proteins which possess a closed (by a cysteine or amide bridge) loop, which is intersected by at least one free end of a backbond chain. More precisely, such a free end intersects a surface of minimal area (constructed theoretically) spanned on a covalently closed loop. Such a surface of minimal area resembles a soap bubble spanned on a closed loop, and its shape can be determined with help of mathematical methods based on triangulation techniques. A surface constructed in this way is referred to as the minimal surface. Analysis based on such surfaces, including tools to determine the length of a terminus (i.e. the number of $C\alpha$ atoms) intersecting the loop, enables identification of a lasso type and its depth. In particular one can determine whether a lasso is shallow (and so it could be untied by thermal fluctuations). In addition, complicated configurations of the backbond chain (e.g. in cases it leads to self-intersections of the minimal surface), are characterized also by a barometric plot. It should also be stressed, that identification of lassos by a "naked eye" is often more difficult than in case of knots, and even impossible in case of larger loops. The model constructed in my work is (to my best knowledge) the first ever tool that enables identification of lasso configurations.

One should note, that identification of lasso structures was motivated by my work on a protein called leptin, where I discovered for the first time a simple lasso configuration [H11], and in the

following work [H4] I constructed a prototype model of a minimal surface to analyze other four proteins.

II.D Classification of proteins with lassos

One of the main results of this work is the identification and characterization of lassos in proteins. Using the model of minimal surfaces described above, and analyzing a representative set of proteins with cysteine bridges (with sequential similarity not larger than 35%) I showed, that 18% of those proteins – which is quite a large number – possess lassos [H1]. I classified various discovered configurations, shown schematically in figure 2, into the following motives:

- single lasso L_1 – the minimal surface spanned on a protein loop is crossed once by one tail (protein terminus);
- double lasso L_2 – after intersecting the minimal surface, the same tail of a protein intersects it again in the opposite direction (so that the loop is crossed twice);
- triple lasso L_3 – a configuration similar to L_2, however the same terminus intersects the loop once more, in the opposite direction (in total the loop is intersected three times);
- supercoiled LS_i – a surface of a loop is intersected at least twice in the same direction (which resembles supercoiled DNA), the index *i* denotes the number of intersections of a loop;
- two-sided lasso LL_{ij} – minimal surface is crossed at least once by each terminus of a protein, indices *i,j* denote the number of intersections by N-terminus and C-terminus.

Figure 2: *Lasso configurations identified in my work (from left to right): L_1 (a single lasso), L_2 (double lasso), L_3 (triple lasso);* LS_i *(supercoiled lasos, index i denotes the number of intersections of the loop, in this example i=2) and LL_{ij} (two-sided lasso, in this example each terminus is threaded once through the loop, so that i=1 and j=1).*

Independently of those results, based on a thorough analysis of how often a given lasso type is detected for a particular secondary structure, a type in the CATH classification, biological function, and the source organism, I found several interesting correlations [H1]. In particular I showed that lasso motifs are most often formed by β hairpins (L_2 motifs appears in 95% situations).

Presentation and classification of lasso structures has been made available online, as the first ever internet server and database *LassoProt*, ***http://LassoProt.cent.uw.edu.pl*** [H1]. The database part of *LassoProt* assembles all entangled structures deposited in PDB (224282 chains) with a loop closed by a cysteine, amide, ester, or tioester bridge. The server part of *LassoProt* can be used for analysis of entanglement of a single chain, or a series of configurations, determined e.g. from molecular dynamics simulations; a user is also provided with a graphical representation of topological changes in time.

LassoProt database and server can be used in multiple ways. In biophysics they can be used e.g. to analyze mechanisms of lasso formation, or to identify new reaction coordinates to describe knotting processes. In biology *LassoProt* may be used to analyze the evolution of protein structures (*LassoProt* database contains all homological sequences, together with information about their

entanglement). In biochemistry it can be used to engineer new structures (e.g. in case of point-like mutations introduced to improve stability, the server could verify whether the entanglement of resulting motifs is correct). The server can be also useful in searching for more complex structures, e.g. links in (bio)polymers. Possible applications of *LassoProt* in other fields are presented in its webpage. It should be noted, that *LassoProt* webpage has been visited more than 7000 times by users from 24 countries in first 3 months after its launch.

III Evolution of topological motifs, geometric configuration and the role of entanglement in proteins

In this chapter, in section **A** I will discuss a surprising evolutional behavior of entangled motifs in proteins, and in section **B** I will explain their role in folding mechanisms. I will then discuss how entanglement affects function of proteins with knots and lassos, respectively in sections **C** and **D**.

III.A Surprising evolutional behavior of entangled motifs

Knot fingerprint (introduced in section II.A above) is a characteristic of a protein, which enables finding correlations between its sequence, structure, functions, and evolution. From a review of fingerprints of protein structures (deposited in PDB) one can deduce, that proteins with very low sequential similarity may have the same entanglement motif. In [H12] I showed, that proteins separated by more than a billion years of evolution possess the same topological motif. For example structures from UCH family represented by human proteins, yeasts and P. falciparum plasmodium cells (protozoan parasite), whose sequential similarity is smaller than 25%, possess the same topological motif $K5_2,3_1,3_1$ (i.e. their subchains contain 5_2 knot and two different knots of type 3_1). As another example, a group of membrane proteins, which are in general responsible for transport of ions, have sequential similarity of the order of 9% and the same motif $S3_1,4_1,3_1$. Because the mechanism of folding proteins to entangled structure is slower and less effective than for unknotted proteins of the same length and similar sequence, preservation of the shape of the entanglement matrix implies a positive role of proteins with nontrivial topology for their hosting organisms. This relation suggests, that knotted motifs survived evolutional selection, because their topology is of advantage for proteins' functions or conditions of their environment. This statement is crucial for this research area.

Before we discuss a biological role of entanglement we should note, that according to the data presented in [H3,H12] even the simplest knot type is strongly conserved in the whole clan (i.e. a group of proteins with sequential or structural similarity, or similar HMM (hidden Markov model) profile available in the Pfam database). The results of my research show that proteins with P-value around 0.04 still possess the same topology, while it is usually assumed, that only for $P<0.001$ structures are regarded as similar. Proteins from N-acetylornithine transcarbamylase (ATCases) and ornithine transcarbamylase (OTCases) families are the only known exceptions, however in my opinion this exception only confirms a general rule. Exceptional behavior of knot motifs suggests, that topology is the fifth, and a very important element of spatial configurations of proteins. Taking entangled motifs into account should largely improve protein structure predictions (e.g. in the CASP competition), especially when sequential similarity of various structures is very low, however there are indications that they have origin in a common family.

III.B Impact of geometric configuration on the process of knot formation

Matrix model (introduced in section II.A), which characterizes entanglements in proteins [H12], can be also used to identify relations between geometric configuration, biological function, and mechanisms of knot formation. Even though explicit correlations between some particular sequences of amino acids and entanglement have not been found to date, in [H9] I showed that borders of a knot (as determined by a matrix model) are correlated with a strong sequential preservation of glycine (one of the weakest evolutionally preserved amino acids), alanine and proline. These regions (i.e. borders of a knot) correspond to hinge-like locations, which I identified (theoretically and in simulations in coarse grained models) as crucial for knot formation [H16,H17,H18].

Matrix model also provides important information necessary for further analysis of evolution of entangled proteins. One possible approach is a method on the border of bioinformatics and statistical physics, which characterizes coevolution of pairs of amino acids, and which I have been developing recently. In this approach a matrix model identifies regions in the protein structure where topology changes may happen, and which should be formed as a result of coevolution of pairs of amino acids (24). Analysis of these regions should also lead to identification of pairs of amino acids (not just single amino acids [H12]) responsible for preservation of nontrivial topology, and should help to identify physical interactions responsible for a folding process. I will discuss mechanisms of folding of proteins with knots in chapter IV, sections A and B.

III.C Role of knots in proteins

One of the key issues concerning entangled structures is how they affect the function of proteins. Correlations between nontrivial topology and biological functions or (chemical, thermal) conditions in which a protein can perform its function, can be identified in a few examples discussed in [H8,H11-H12,H15-H16,H18-H19].

Analysis of data from the *KnotProt* database described above shows, that 91% of proteins with nontrivial topology have enzymatic functions [H12]. Moreover it turns out, that a location of a knot is correlated with an active center (as I have shown in detailed analysis of three families of knotted proteins). In connection with the fact (mentioned earlier), that protein's topology is preserved stronger than its sequence and tertiary structure, we can conclude, that the location of an active center and its signal path are strongly influenced by topological constraints. This statement is illustrated e.g. in [H12], where I showed, that slipknot configuration plays a role of a "belt" which connects transmembrane helices – the structure stabilizes a helical channel during active ion transport, which requires large conformational changes.

Another important role of knots in proteins can be identified by the analysis of correlations between their appearance and properties of surrounding environment. According to details presented in *KnotProt* database [H12], most entangled proteins function in extreme conditions, e.g. in hot springs, or at very high or very low pH. Such circumstances suggest, that a role of non-trivial topology may be to prevent proteins, and in particular their active centers, from thermal or mechanical degradation. In [H19] I showed that knotted proteins have higher stability during thermal or mechanical degradation, as compared to unknotted proteins with the same fold, structure, and high sequential similarity. In [H15,H19] I also showed that, unexpectedly, during mechanical stretching a knot tightens in the region of a native state, and in collaboration with experimentalists I showed that a similar phenomenon takes place during chemical denaturation [H8]. The existence of knots in a denatured state was postulated also in a few other experimental works (25).

The above considerations imply, that experimental untying of a knot in a protein – one of the most important challenges in this field in last years – appears to be impossible without destroying the whole structure of a protein. This surprising result is also related to proteins' lifetime. On one hand we know, that a probability of spontaneous tying or untying of a knot is very low. On the other hand, a probability of correct folding of a protein from the state which is (initially) knotted is equal to almost 100% [H17]. This implies that the presence of a knot prolongs proteins' lifetime in an unfavorable environment, because partial unfolding of its structure upon thermal or chemical denaturation takes a protein away from its native state only for short time. In conclusion one can postulate, that difficulties in the folding process are compensated by a prolongation of time period in which a protein is active.

Independently of the above conclusions, in chapter V I will discuss further properties of the process of mechanical unfolding of proteins with knots and slipknots.

III.D Role of lassos in proteins

I characterized the role of lassos in proteins in publications [H1,H4,H13]. In [H1], based on biological classification, I showed that proteins with lassos appear more often in viruses, plants and fungi than in other kingdoms. Moreover, contrary to proteins with knots and slipknots, only 38% of lasso proteins have enzymatic functions. I also showed that various functions of proteins are typical for a given type of a lasso. Furthermore, I identified proteins, whose biological function may be supported by lasso motifs; in particular:

- in anhydrases, which possess very shallow knots (which can be untied by thermal fluctuations), a lasso motif supports stable conformation of an active center,
- in a family of proteins with RNase activity (1 RNase H from hyperthermophilic archeons of organisms which live in temperature above 100 °C), a lasso of L_3 type seems to be responsible for a very high thermal stability of those proteins, which is achieved by imposing constraints on their unstructured active centers. A similar role may be played by a lasso motif in unstructured proteins with L_1 topology, which constitute around 5% of proteins.

These and other examples are presented in the *LassoProt* ***http://LassoProt.cent.uw.edu.pl*** [H1]. Among important examples, I confirmed a biological role of a lasso configuration in leptin, which is the first protein where I identified a lasso motif [H13]. Leptin plays a crucial role in food and energy consumption. This hormone acts on receptors in the arcuate nucleus of the hypothalamus to regulate appetite (i.e. it determines the excitation or inhibition of appetite, *Nat Clin Pract Endocrinol Metab* ***2*** *(6): 318–27*). In [H13], based on experimental and theoretical results, I showed that a cysteine bridge (responsible for the existence of the lasso motif) plays an important role in the receptor binding, and therefore it also plays a role in activation of the biological function by turning off vibrations in the area of receptor binding, even though this area is not directly connected with a cysteine bridge. Moreover, having analyzed proteins with the same fold and lasso motif L_1, I showed that changes in the dynamics of the native state (in the vicinity of an active center) are determined by and depend on the location and the size of a loop [H4].

Independently of those results, in [H1] I showed that lasso motifs exist in miniproteins (27) that, among others, are used for pharmacological purposes. The methods which I developed for identification of lasso motifs provide new tools to analyze these miniproteins and should be useful in engineering new therapeutic solutions (not only in proteins with a cysteine bridge).

In summary, in [H1,H4,H13] I showed that a lasso structure is an additional element that controls function and energetic landscape of proteins (described in chapter IV). This structure should be taken into account e.g. upon engineering of new drugs, in particular for proteins with a four helix bundle fold [H4], or in miniproteins [H1]. It should be stressed that, contrary to knotted proteins, a single cysteine mutation may cause a topology change in proteins with lassos. This property may be used to steer biological activity of a protein, via topology change induced by a single mutation, or by changes from oxidation to reductive conditions.

IV The free energy landscape – formation of knots, slipknots and lassos

In this section I will show that the process of folding of entangled proteins, in contrast to proteins with trivial topology, reminds a chemical reaction with an unambiguous sequence of events determined by the conformation of transition states. The corresponding free energy landscape is not perfectly smooth, and for most proteins it is still not possible to determine it by means of theoretical models (numerical simulations). Moreover I will show that commonly used reaction coordinates (e.g. the number of native contacts) cannot be used to identify transition states in the free energy landscape of knotted proteins. To determine the energy landscape of entangled proteins I performed molecular dynamics simulations with the coarse-grained Structure Based Models (SBM) with different graining, and with explicit solvent all atom models with different force fields (Desmond - Schrödinger).

In this section I discuss folding mechanisms of entangled proteins: proteins with deep knots and slipknots in section **A**; proteins with rather shallow knots, including a detailed characterization and interpretation of their energy landscape, in section **B**; proteins with lassos, in section **C**.

IV.A Folding mechanisms of proteins with deep knots

Theoretical models of protein folding, which use the information about the native state of proteins, are constructed based on the assumption of the minimal frustration of the free energy landscape. Such models are one of basic tools for understanding how relatively weak molecular interactions lead to a smooth and cooperative protein folding (11,12). Because mechanisms of protein folding, as well as functional motions, are governed by the same energy landscape, it follows that this energy landscape should enable a characterization of relations between structure, folding, and function of proteins. Theoretical studies based on such assumptions, conducted in recent years, revealed a number of new mechanisms responsible for regulation of biological functions of proteins (including molecular machines, e.g. rybosomes), without a need to conduct expensive and difficult experiments (12). However it turns out, that in case of knotted proteins such an approach must be modified.

In the standard approach, the funnel energy landscape theory assumes that proteins can reach their native conformations via a set of routes. The number of possible conformations decreases upon approaching the native state. Folding starts with a formation of nucleation sites, which are gradually expanded by the formation of other secondary structures, which finally condense to from a unique, native three-dimensional conformation of a protein. The driving force to fold a protein comes from native attractive interactions, which are responsible for a smooth downhill route towards the native state. However these statements are not consistent with behavior of knotted proteins, and their description requires modification of the above mentioned theory. In my works I postulated, that such a modification should involve a presence of topological barriers in the free energy landscape, whose overcoming is possible in the presence of some specific interactions in proteins with deep knots. As

will be explained in what follows, in my work I identified such barriers and described their properties.

It is also worth mentioning that initially, in experimental studies of knotted proteins, it was incorrectly assumed that such proteins easily fold and unfold. However today we are still unable to untie proteins experimentally, and we do not have tools to track changes of topology of proteins during folding and unfolding processes. For this reason the theoretical approach is the only tool that can be used to understand relations between entanglement and conformation, and it is with this tool that we managed to understand folding mechanisms of knotted proteins.

The most important results of my studies of folding processes of proteins with deep knots are as follows. In publications [H10,H14,H16,H17] I showed that knotted proteins (from three groups of with different folds and functions) can get tied successfully in structure based coarse grained models (contrary to earlier expectations (28)). In [H17] I showed that native contacts are sufficient to tie proteins YibK and YbeA, from SPOUT family, which do possess deep knots; however their knotting probability is very low, and the funnel landscape is restricted to only one possible route, (in contrary to most proteins with trivial topology). For those proteins the folding route consists of five main steps, where the first three conformational changes correspond to Reidemeister moves I and II (well known in knot theory). Knotting arises after the native twisted loop is formed (in the fourth step), through which the shorter terminal of the protein in non-native conformation is threaded (in the last, fifth step). The analysis of the distribution of folding times shows that threading the tail (and forming the knot) is the main rate limiting step on the folding route. The obstacles which make this last step difficult to overcome I called the topological barrier.

Contrary to knots formed spontaneously and in random positions in polymers, knots in proteins are always formed in the same place and at the last step of the folding pathway. Even though I found that the probability to tie proteins is very low, the above results, found 2009 in [H17], were very important in this area, because in contrast to earlier interpretation of experimental results I showed, that folding times of knotted proteins are surprisingly long, and are determined by the moment of knotting. In 2012 my theoretical studies were confirmed experimentally by Prof. S. Jackson (Cambridge University), who showed that folding times of knotted proteins are around 20 minutes, rather than seconds. In those experiments it was also shown that short folding processes, of the order of a few seconds, are only possible if denatured protein backbond still remains knotted, as I also argued in [H17].

In the publication [H11] in the theoretical SBM model I investigated the free energy landscape of the first artificially engineered (by domain fusion) knotted protein. I showed that unlike to previously studied knotted proteins of similar length [H17], this protein can omit topological traps and tie and untie reversibly. To understand how evolution might further select the geometry and stiffness of the threading region, I analyzed several mutations in this protein. I showed that similarly to the wild-type protein, all its mutants posses robust and reversible folding routes. As also observed experimentally, my simulations show that the knotted protein fold about ten times slower than the unknotted construct with “the same” contact map. Moreover I showed that numerical simulations with the SBM model reproduce experimentally observed rollover in the folding limbs of chevron plots, and with comprehensive analysis of numerical data I explained that this rollover (at low temperature) could be a consequence of a non-diffusive dynamics of the knotted protein. Additionally I showed that successful folding of the knotted protein is restricted to a narrow range of temperatures as compared to the unknotted protein. Recently, in 2015, these results were confirmed in experiments conducted in the group of Prof. S. Jackson.

In [H14] I showed that the energy landscape of the smallest knotted protein is also minimally frustrated. Historically it was the first protein whose thermodynamics properties, together with topological changes, were determined by means of The Weighted Histogram Analysis Method (WHAM). In [H14] I showed that independently of the coarse graining of the model (e.g. with

amino acids represented only by *Cα* atoms, or all heavy atoms), a knot is always formed in the last phase of folding, after overcoming the topological barrier. Moreover I showed that shallow knots in a denatured state are very rare, and deep ones are not observed. I also showed that folding probability and a choice of the protein's terminus and geometry (trivial or slipknot-like), used to cross the topological barrier, are determined by a depth of a knot. I confirmed these statements also in [H5], analyzing (in SBM model) in addition a folding process of a protein with the same fold but different knot depth.

In order to verify results obtained in SBM models, I also performed folding simulations of the smallest protein in the CABS model, constructed by Prof. Koliński. The hamiltonian of this model is based on statistical potentials. I found that a probability of reaching a knotted state in this model is 20%, while it is more than 90% in SBM models used and improved by me.

To sum up, in my studies I characterized the process of protein folding, taking into account the presence of knots. The analysis of kinetic and thermodynamic processes of knotted proteins shows, that knots are formed after a native loop is formed, and then a shorter end of a protein is threaded through this loop, either in trivial or slipknot-like conformation. This simple process – however not very intuitive, when compared to how knots are formed in polymers – is easy to explain, when one observes that all knots identified in proteins belong to a family of so called "twist knots" [H12]. The process of formation of twist knots was already postulated in [10,31], however at that time much less knotted proteins were known, and in particular a protein with 6_1 knot was not yet discovered. In twist knots a reduction of a single crossing results in their untying, and therefore 3_1, 4_1, 5_2, and 6_1 knots may be formed after appropriately twisted look is created, and one terminus is threaded through this loop. In [H16] I analyzed a folding process of a protein with a complicated 6_1 knot, and confirmed that this hypothesis is true (even though a probability of knot formation in this case is smaller than in SpoUT case). One should also note, that last experiments of Prof. S. Jackson (Cambridge University) show, that folding times of proteins from SpoUT family may be reduced by an order of magnitude by chaperons – it may explain, why in my theoretical analysis, which does not take the presence of chaperons into account, knotting probabilities are small. In some of my other publications, which are not part of this work, I showed that the knot formation should take similar form also in presence of chaperons. Yet another argument for the correctness of the knot formation process described above is the fact, that to date no protein with 5_1 knot (which is not a twist knot, and its formation would require threading one end through two different loops) has been found. This confirms that knots in proteins are formed after threading one knot terminus through one (possibly twisted several times) native loop.

IV.B Detailed analysis of folding of the smallest knotted protein

With coarse grained SBM models one can analyze thermodynamics properties of a given system, if it is characterized by a smooth energy landscape. However, as I showed above, crossing of the topological barrier requires dense, non-native packing of a protein structure, which is accompanied by appearance of non-native contacts. These non-native contacts are taken into account in SBM models by the excluded volume effect, which very likely slows down or blocks the folding process. In [H9], performing all atom simulations in explicit solvent (on the Anton, D. Shaw supercomputer with Desmond program), I found that a spontaneous self-tying of the smallest protein is possible, via an intermediate slipknot configuration. I also showed, that the same interactions provide driving force for folding in both types of models (SBM with implicit solvent and all atom with explicit solvent). A thorough analysis of the folding mechanism showed, that non-native contacts actively participate in threading one terminus through a native loop (contrary to the postulate, that they are important only in the initial phase of this process (28)). Therefore, combining results of [H9] and [H14], we see that SBM models can be used to analyze knotted proteins, at least those with a similar fold.

Moreover, in [H2] I showed (using Bayesian analysis and constructing conditional probabilities between equilibrium states and protein conformation around maxima in the energy landscape $F(Q)$), that the number of native contacts Q, commonly used in the analysis of proteins with non-trivial topology (29), does not describe properly the process of overcoming the topological barrier. Moreover variational methods, which I used to optimize a choice of the coordinate system, are not very useful (in contrast to globular proteins (18)) in identification of protein conformation in the transition state. This is a surprising result, because a growing number of papers shows, that native contacts determine protein folding also in all atom simulations, explicit solvent (18). In summary, the results of my work show, that while native contacts are sufficient to overcome the topological barrier, the number of native contacts does not provide a good coordinate to characterize this process. These are important conclusions, which indicate possible directions for future research.

Another important issue, which affects the smoothness of the energy landscape in SBM models, is the number and the way of identification of native contacts. In [H2,H4] I characterized the minimal number of physical contacts necessary to overcoming a topological barrier and folding a protein to its native state. A possibility of engineering a minimally frustrated landscape of interactions for the smallest knotted proteins suggests, that an analogous map may exist for proteins with deep knots (e.g. proteins from SpoUT family described above) – however currently we are not able to determine such a map. A minimally frustrated sequence (with a smooth energy landscape) also in case of proteins with a deep knot may be a result of the evolution process (i.e. a selection of a specific sequence). This could ensure an effective folding process of knotted proteins, while fast denaturation (being a consequence of a small number of stabilizing native interactions) would be shielded by topological constraints.

IV. C Energy landscape of proteins with lassos

Proteins with lassos constitute another important group of entangled biomolecules. Analysis of these proteins in oxidation and reduced states (corresponding respectively to the presence or absence of a lasso) is an important tool, which also sheds light on some properties of knotted proteins.

In [H4,H13], in the analysis of proteins with the same fold with either (in five cases) a single lasso of L_1 type, or trivial (in three cases), I showed, that the energy landscape of lasso proteins is minimally frustrated. Native contacts turn out to be sufficient to overcome a topological barrier, which in case of lassos corresponds to threading one terminus of a protein through a closed covalent loop. Having constructed various tools to analyze lasso geometry, I showed that in the process of overcoming of a topological barrier, if the length of the closed loop grows, a conformation of the threaded terminus changes from a slipknot-like (curved) to a straight one. Crossing the topological barrier is the process that determines time and probability of folding.

Thorough analysis (using SBM model with $C\alpha$ and all atom representations) of a single protein (leptin) showed [H4], that energy landscapes and folding pathways of proteins in reduced and oxidation states are surprisingly similar to each other. Furthermore, comparison of experimental and theoretical results showed differences in the size of the free energy barrier (as the function of the number of contacts) in three systems: reduced one without a possibility of forming a bridge during the folding/unfolding process, reduced one with a possibility to form such a bridge, and in oxidation state (when the bridge is present during the whole simulation). This difference probably follows from the fact, that the number of native contacts Q cannot be used as a reaction coordinate to describe energy landscape of proteins with nontrivial topology. This statement is consistent with results of [H2], where I also showed (after analysis of folding of a knotted protein), that the number of contacts Q indeed cannot be used to determine a transition state.

To sum up, the above results show that for lasso proteins the evolution also produced a minimally frustrated sequence, and in consequence a smooth energy landscape. Moreover folding is a complex process, which can be analyzed by means of SBM models, however a proper description

of transition states requires using new coordinates, other than the number of native contacts Q or the mean standard deviation (RMSD). A size of the minimal surface considered in [H1] may play a role of such a new coordinate.

V The free energy landscape – unfolding and untying of knots and slipknots

In previous chapters I discussed properties of the energy landscape from the point of view of kinetics and thermodynamics of protein folding. In this chapter I will characterize the free energy landscape from the viewpoint of unfolding pathways and conformations of proteins in the denatured state. Analyzing the results of computer simulations and experimental data I will show, that untying of proteins is a complex process, still not realized experimentally [H8]. However, on the basis of theoretical mechanical stretching, in this section I will present a theoretical model of knot untying [H15]. In addition, I will show that metastable states with high mechanical resistance appear during stretching of proteins with slipknots [H18].

V.A Knotted proteins in denatured state and a model to untie them

Thermal and chemical denaturation are important processes, which connect a unique biologically active conformation of a protein with its random coil state. To study those processes the condition of the reversible denaturation must be met. In case of proteins with trivial topology, the aggregation is the primary phenomenon limiting reversible denaturation.

Based on comprehensive experimental and theoretical studies, in [H12] I showed that knot untying is a process that blocks a reversible denaturation. Analysis of kinetic data obtained in computer simulations shows that unfolding process consists of at least three steps. In the last step of this process a knot always tightens around its native location. Untying time is at least one order of magnitude slower than unfolding; experimentally observed hysteresis loop (during chemical denaturation) is a consequence of the existence of two possible refolding pathways, one from random but knotted conformation of a protein, and the second one (however quite rare) from random and unknotted conformation. These results clearly show that a unique feature of knotted proteins is (unfortunately) the existence of knots in their backbonds in a denatured state. This property has been recently confirmed in independent experimental studies conducted by Prof. S. Jackson (Cambridge University).

Finally let me summarize the results of [H15]. The objective of this work was to construct a theoretical model to estimate a probability of mechanical untying (upon stretching) of proteins from SpoUT family based. The main advantage of the method of mechanical stretching is a reduction of the free energy barrier (if the stretching direction follows a natural, thermodynamic reaction coordinate), and consequently more accurate probing of the energy landscape in order to identify all transition states. In [H15], based on simulations of stretching of proteins in various directions in the SBM model, with different choices of amino acid pairs by which a protein is stretched, and analyzing different temperatures stretching speeds, I determined a probability of knot untying. Moreover I found a general formula for the probability of untying of proteins from SpoUT family by mechanical manipulation. The model formulated in [H15] is crucial for future experimental work on knotted proteins based on AFM and optical tweezers, and in particular for developing effective experimental techniques of knot untying.

V.B Unique metastable properties of proteins with slipknots

From a mathematical point of view a topology of a slipknot is trivial, so that one could expect, that pulling its ends will transform it into a straight segment. In [H18] I conducted a theoretical analysis (in SBM model) of stretching of proteins with slipknots, with different speeds and forces. I showed that details of such stretching processes depend on geometry and elastic properties of proteins, a direction of stretching, a magnitude and attachment of pulling forces, and a pulling speed, and in particular may lead to a jammed slipknot conformation. When a pulling force or speed is high, the unfolding process consists of two stages, with an intermediate state corresponding to a metastable geometry of a protein. I showed that the unfolding process through such metastable states is described by the Bell equation. Having analyzed bending energy of a protein chain and the mutual friction of peptide chains between two loops in a slipknot, I formulated a theoretical model, which describes conditions necessary for the emergence of the above mentioned metastable states. It should be emphasized that these results have been already confirmed experimentally in (30) by a group of Prof. Hongbin Li (University of British Columbia, Canada). Even though this theoretical model was formulated for globular proteins, it is also a valuable tool to study membrane proteins, which quite often possess slipknots.

Bibliography

[1] M.L. Mansfield, *Are there knots in proteins?*, Nat Struct Biol (1994) 1:213–214.
[2] W.R. Taylor, *A deeply knotted protein structure and how it might fold*, Nature (2000) 406:916–919.
[3] R.C. Lua, A.Y. Grosberg, *Statistics of knots, geometry of conformations, and evolution of proteins*, PLoS Comput Biol (2006) 2:e45.
[4] P. Virnau, L.A. Mirny, M. Kardar, *Intricate knots in proteins: Function and evolution*, PLoS Comput Biol (2006) 2:e122.
[5] N.P. King, E.O. Yeates, T.O. Yeates, *Identification of rare slipknots in proteins and their implications for stability and folding*, J Mol Biol (2007) 373:153–166.
[6] K. Shimokawa, K. Ishihara, I. Grainge, D. Sherratt, M. Vazquez, *FtsK-dependent XerCD-dif recombination unlinks replication catenanes in a stepwise manner*, PNAS (2013)110, 52: 20906-20911.
[7] P.Y. Chou, G.D. Fasman, *Prediction of protein conformation*, Biochemistry (1974) 13, 2: 222–245.
[8] K.L. Tkaczuk, S. Dunin-Horkawicz, E. Purta, J.M Bujnicki, *Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases*, BMC Bioinformatics (2007) 8:73.
[9] R. Potestio, C. Micheletti, H. Orland, *Knotted vs. Unknotted Proteins: Evidence of Knot- Promoting Loops*, PloS Comput Biol (2010) 6:e1000864.
[10] J. I Sulkowska, *Stretching and folding proteins in coarse grained models,* doktorat (2008).
[11] P.G. Wolynes, *Evolution, energy landscapes and the paradoxes of protein folding,* Biochimie (2014) S0300-9084, 14:00389-7.
[12] P.C. Whitford, J.N. Onuchic, *What protein folding teaches us about biological function and molecular machines*, Current Opinion Struct Bio (2015) 30, 57–62.
[13] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, *Funnels, pathways, and the energy landscape of protein folding: a synthesis*, Proteins (1995) 21, 3:167-95.
[14] G.D. Rose, P.J. Fleming, J.R. Banavar, A. Maritan, *A backbone-based theory of protein folding*, PNAS (2006) 103, 45:16623–16633.
[15] K.A. Dill, *Dominant forces in protein folding*, Biochemistry (1990) 29 (31): 7133–7155.
[16] A. Godzik, A. Kolinski, J. Skolnick, *Topology fingerprint approach to the inverse protein folding problem*, J Mol Biol (1992) 227, 1, 227-238.
[17] C. Levinthal, *Are there pathways for protein folding?,* Journal de Chimie Physique et de Physico-Chimie Biologique (1968) 65: 44-45.

[18] R.B. Best, G. Hummer, W.A. Eaton, *Native contacts determine protein folding mechanisms in atomistic simulations*, PNAS (2013) 110(44): 17874-17879.
[19] E.R. Henry, R.B. Best, W. A. Eaton, *Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations,* PNAS (2013) 110, 44, 17880-17885.
[20] K.C. Neuman, A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*, Nat Methods (2008) 5(6): 491-505.
[21] J.M. Fernandez J.M, H. Li, *Force-clamp spectroscopy monitors the folding trajectory of a single protein*, Science (2004) 12;303(5664):1674-8.
[22] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, K. Schulten, *Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation*, Biophys. J. (1998) 75, 2, 662-671.
[23] J.I. Sułkowska, M. Cieplak, *Stretching to understand proteins - a survey of the protein data bank*, Biophys J. (2008) 94,1: 6-13.
[24] J.I. Sułkowska, F. Marcos, T. Hwa, J. N. Onuchic, *Genomics Aided Structure Prediction (GASP)*, PNAS (2012) 109, 26:10340-5.
[25] A.L. Mallam, J. M. Rogers, S. E. Jackson, *Experimental detection of knotted conformations in denatured proteins*, PNAS (2010) 107: 8189-8194.
[26] N.P. King, A.W. Jacobitz, M.R. Sawaya, L. Goldschmidt, T.O. Yeates, *Structure and folding of a designed knotted protein*, PNAS (2010) 107: 20732-20737.
[27] Yanyan Li, Séverine Zirah, Sylvie Rebuffat, *Lasso Peptides: Bacterial Strategies to Make and Maintain Bioactive Entangled Scaffolds,* (2014) Springer, Oct 21.
[28] S. Wallin, K.B. Zeldovich, E.I. Shakhnovich, *The folding mechanics of a knotted protein*, J Mol Biol (2007) 368, 3: 884-93.
[29] R.B. Best, G. Hummer, *Reaction coordinates and rates from transition paths*, PNAS (2005), 102, 19:6732-7.
[30] Ch. He, G. Lamour, A. Xiao, J Gsponer, H Li, *Mechanically Tightening a Protein Slipknot into a Trefoil Knot*, JACS (2014) *136* (34), 11946-11955.
[31] W.R. Taylor, *Protein knots and fold complexity: Some new twists*, Comput Biol Chem (2007) 31, 151-162.

5. Description of other scientific achievements

A) Grants and research projects

Research grants – 7 as a principal investigator, 2 as a main investigator, 5 as an investigator
Conference grants – 7 as a principal investigator / main organizer, 4 as a partner

14. Project type: **Ideas Plus** (2 615 024 PLN) – for realization of the "ERC Starting grant" project accepted for realization (but not financed due to budget restrictions) by the European Research Council
Project title: **Białka splątane - studium nowych struktur i rozwiązanie ich zagadki**
Starting year of the project: **2016**
Place of realization: **Faculty of Chemistry, Centre of New Technologies, University of Warsaw**
Funding agency: **Ministry of Science and Higher Education (MNiSW)**
Role in the project: **Principal Investigator**

13. Project type: **Early Science** (computer time on supercomputer Okeanos)
Project title: **Influence of substrate binding on dynamics of knotted protein**
Starting year of the project: **2016**
Place of realization: **Faculty of Chemistry, Centre of New Technologies, University of Warsaw**
Funding agency: **ICM, UW**
Role in the project: **Principal Investigator**

12. Project type: **EMBO Small Grant** (8400 Euro)
Project title: **Synteza inhibitorów dla białek z rodziny TrmD**
Starting year of the project: **2016**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **European Molecular Biology Organization**
Role in the project: **Principal Investigator**

11. Project type: **EMBO Installation Grant** (250 000 Euro)
Project title: **Comprehensive analysis of knotted proteins**
Starting year of the project: **2014**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **European Molecular Biology Organization**
Role in the project: **Principal Investigator**

10. Project type: **International, interdisciplinary „Skills/Inter"** (100 000 PLN)
Project title: **Protein with a knot as the Gordian knot**
Starting (and end) year of the project: **2013 – 214**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **Foundation for Polish Science**
Role in the project: **Principal Investigator**

9. Project type: **Sonata Bis (**1 490 000 PLN**)**
Project title: **Influence of knotted structure on function of proteins and protein structure prediction**

Starting year of the project: **2013**
Place of realization: **Faculty of Chemistry, Centre of New Technologies, University of Warsaw**
Funding agency: **National Science Centre, Poland**
Role in the project: **Principal Investigator**

8. Project type: **Grant Homing-Plus** (253 000 PLN)
Project title: **Entanglement in biology – how nature controls the topology of proteins**
Starting (and end) year of the project: **2012 – 2014**
Place of realization: **Faculty of Chemistry, Centre of New Technologies, University of Warsaw in collaboration with University of California San Diego**
Funding agency: **Foundation for Polish Science**
Role in the project: **Principal Investigator**

7. Project type: **Anton Grant PSCA00062P (**computing time **–** D. E. Shaw Research**)**
Project title: **To knot or not to knot: slipknotting in the smallest knotted protein**
Starting (and end) year of the project: **2011-2012**
Place of realization: **University of California San Diego**
Funding: **National Research Council (NRC) at the National Academy of Sciences of USA**
Role in the project: **Principal Investigator**

6. Project type: **Promotor Grant ("Grant promotorski"), nr N202 021 31/0739**
Project title: **Folding and unfolding proteins in structure based model**
Starting (and end) year of the project: **2006 – 2007**
Place of realization: **Institute of Physics, Polish Academy of Sciences**
Funding agency: **Ministry of Science and Higher Education (MNiSW)**
Role in the project: **Principal Investigator**

5. Project type: **International, interdisciplinary, PHY-1212312 - prof. P Jennings**
Project title: **Entanglement in Biology – How Nature Controls the Topology of Proteins**
Starting (and end) year of the project: **2012 – 2015**
Place of realization: **University of California San Diego, USA**
Funding agency: **National Science Foundation of USA**
Role in the project: **Main investigator**

4. Project type: **International, interdisciplinary – prof. E Radwan**
Project title: **Theory and simulations of knotting in physical and biological systems ranging from proteins to glueballs**
Starting (and end) year of the project: **2011 – 2014**
Place of realization: **University of St. Thomas, USA**
Funding agency: **National Science Foundation of USA**
Role in the project: **Main investigator**

3. Project type: **International, interdisciplinary „Skills/Inter" – mgr Wanda Niemyskiej**
Project title: **Bańki mydlane – Czy biologie da sie schwytac na matematyczne lasso?**
Starting (and end) year of the project: **2015 – 2016**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **Foundation for Polish Science**
Role in the project: **Main investigator**

2. Project type: **International, interdisciplinary, PHY- 1308264 – prof. Jose Onuchic**
Project title: **Sharing the Energy Landscape for Folding an Function: from Proteins to Biomolecular Machines**
Starting (and end) year of the project: **2011 – 2016**
Place of realization: **University of California Sand Diego, Rice University (USA)**
Funding agency: **National Science Foundation of USA**
Role in the project: **Main investigator**

1. Project type: **International, interdisciplinary „Skills/Inter" – dr. hab. Piotra Sułkowskigo**
Project title: **On topology, interacting RNA, and quantum physics**
Starting (and end) year of the project: **2012 – 2013**
Place of realization: **Faculty of Physics, University of Warsaw**
Funding agency: **Foundation for Polish Science**
Role in the project: **Main investigator**

<u>Conference grants</u> – 7 as a principal investigator / main organizer, 4 as a partner

11. Project type: **Conference grant**
Project title: **The Geometry and Topology of Knotting and Entanglement in Proteins**
Year of the conference: **2017**
Place of realization: **Oaxaca, Mexico**
Funding agency: **Banff International Research Station (BIRS), Banff, Canada**
Role in the project: **partner**

10. Project type: **Networking Grant, EMBO** (5 000 Euro)
Project title: **VSSSB2016 symposium**
Year of the conference: **2016**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **EMBO**
Role in the project: **Principal Investigator / main organizer**

9. Project type: **Networking Grant, EMBO** (4 800 Euro)
Project title: **EMBO workshop on Computational Biology II**
Year of the conference: **2016**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **EMBO**
Role in the project: **Principal Investigator / main organizer**

8. Project type: **Networking Grant, EMBO** (15 000 Euro)
Project title: **YSF, EMBO**
Year of the conference: **2015**
Place of realization: **Centre of New Technologies, University of Warsaw**
Funding agency: **EMBO**
Role in the project: **Principal Investigator / main organizer**

7. Project type: **Small grant, Visegrad** (6 000 Euro)
Project title: **3rd Summer School in Molecular Biophysics and Systems Biology**
Year of the conference: **2015**
Place of realization: **Nover Hrady, Czech Republic**

Funding agency: **Visegrad fund**
Role in the project: **partner**

6. Project type: **Networking Grant, EMBO** (3 000 Euro)
Project title: **EMBO workshop on Computational Biology** I
Year of the conference: **2015**
Place of realization: **Biebrza, Poland**
Funding agency: **EMBO**
Role in the project: **Principal Investigator / main organizer**

5. Project type: **Small grant, Visegrad** (6 000 Euro)
Project title: **2nd Summer School in Molecular Biophysics and Systems Biology**
Year of the conference: **2014**
Place of realization: **Nover Hrady, Czech Republic**
Funding agency: **Visegrad fund**
Role in the project: **partner**

4. Project type: **Networking Grant, EMBO** (3 000 Euro)
Project title: **Biophysical Society Meeting**
Year of the conference: **2014**
Place of realization: **Faculty of Chemistry, University of Warsaw**
Funding agency: **EMBO**
Role in the project: **Principal Investigator / main organizer**

3. Project type: **DUN** 782/P-DUN/2014 – (47 000 PLN)
Project title: **Biophysical Society Meeting**
Year of the conference: **2014**
Place of realization: **Faculty of Chemistry, University of Warsaw**
Funding agency: **Ministry of Science and Higher Education (MNiSW)**
Role in the project: **Principal Investigator / main organizer, in collaboration with Polish Biophysical Society**

2. Project type: **Biophysical Society grant** (10 000 USA)
Project title: **Biophysical Society Meeting, Warsaw, Poland**
Year of the conference: **2013**
Place of realization: **Faculty of Chemistry, University of Warsaw**
Funding agency: **Biophysical Society, USA**
Role in the project: **Principal Investigator / main organizer**

1. Project type: **Conference grant**
Project title: **Entanglement in biology; how nature controls the topology of proteins and DNA**
Year of the conference: **2012**
Place of realization: **Banff, Canada**
Funding agency: **Banff International Research Station (BIRS), Banff, Canada**
Role in the project: **Principal Investigator / main organizer**

B) Awards

5. **Grant Ideas Plus** (2 615 024 PLN) – MNiSW grant for realization of the "ERC Starting grant" project (which obtained the highest ERC score A and was accepted for realization by the European Research Council, but was not financed due to budget restrictions)

4. Name of the award: **Outstanding Women in Science**
Year of the award: **2016**
Awarding/funding institution: **Foundation of Robert Bosch, nominated by FNP**
To whom the award is given: **"for outstanding young researchers"**

3. Name of the award: **CBSB14 Outstanding Young Researcher Award**
Year of the award: **2014**
Awarding/funding institution: **Computational Biophysical on Systems Biology Group**
To whom the award is given: **"for outstanding young researchers"**

2. Name of the award: **Installation Grant**
Year of the award: **2013**
Awarding/funding institution: **European Molecular Biology Organization (EMBO)**
To whom the award is given: **"for outstanding young researchers"**

1. Name of the award: **Award for the best PhD thesis in IFPAN**
Year of the award: **2008**
Awarding/funding institution: **Institute of Physics, Polish Academy of Sciences (IFPAN)**

C) Bibliometric data

Total Impact Factor: **183.86**

Total number of citations according to Web of Science data base (however, this data base does not find all citations to my publications): **833**

Total number of citations according to Web of Science data base (however, this data base does not find all citations to my publications) without self-citations: **706**

Total number of citations, (it should be noted, that the Web of Science database does not properly find all citations of my publications because of the Polish letter "ł" and change of the family name; including not cited papers based on the Google scholar database is: **1113**

Hirsch index according to the Web of Science (WoS) database: **16**

The number of all publications: **36**
The number of publications as a „corresponding author": **8**
The number of publications as the first author: **16**

D) A list and description of other publications

Publications in the field of protein structures prediction

1. J. I Sułkowska*, F. Marcos*, T. Hwa, J.N. Onuchic,
"*Genomics Aided Structure Prediction (GASP)*",
Proc. Natl. Acad. Sci. USA (2012), 109(26): 10340-5.

Publications characterizing the energy landscape of proteins with a trivial topology based on the mechanical manipulation

2. Li Sun, Jeffrey K. Noel, J. I. Sulkowska, Herbert Levine, José N. Onuchic,
"Connecting Thermal and Mechanical Protein (Un)folding Landscapes",
Biophys J. 1(2014) 6, 107(12):2941-52.

3. A. Valbuena, J. Oroz, R. Hervas, A. M. Vera, D. Rodrigues, A. Menedez, J. I. Sułkowska, M. Cieplak and M. Carrion-Vazquez,
"On the remarkable mechanostability of scaffoldins and the mechanical clamp motif",
Proc. Natl. Acad. Sci. (2009) 106, 13791.

4. M. Sikora*, J. I. Sułkowska* and M. Cieplak,
"Mechanical strength of 17 134 model proteins and cysteine slipknots motive",
PLoS Comput. Biol. (2009) 5, e1000547.

5. M. Sikora, J.I. Sułkowska, B.S. Witkowski, M. Cieplak,
"BSDB: the biomolecule stretching database",
Nucleic Acids Res. (2011) 39:D443-50. doi: 10.1093/nar/gkq851.

6. M. Cieplak, J. I. Sułkowska,
"Tests of the Structure-Based Models of Proteins",
Act. Phys Polonica A (2009) 115, 441.

7. J. I. Sułkowska, A. Kloczkowski, T. Z. Sen, M. Cieplak and R. L. Jernigan,
"Predicting the Order in Which Contacts Are Broken during Single Molecule Protein Stretching Experiments",
Proteins: Structure, Function, and Bioinformatics (2008) 71, 45-60. DOI: 10.1002/prot.21652

8. J. I. Sułkowska, M. Cieplak,
"Selection of optimal variants of Go-like models of proteins through studies of stretching",
Biophys. J. (2008) 95, 3174.

9. J. I. Sułkowska, M. Cieplak,
"Mechanical Stretching of proteins - A theoretical survey of the Protein Data Bank",
J. Phys. Cond. Mat. (2007) 19, 283201.

10. J. I. Sułkowska, P. Sułkowski, P. Szymczak, M. Cieplak,
"*Tightening of knots in proteins*",
Phys. Rev. Lett. (2008) 100, 058106.

11. M. Cieplak, J. I. Sułkowska
"Thermal unfolding of proteins",
J. Chem. Phys. (2005) 123, 194908.

12. J. I. Kwiecińska (moje nazwisko panieńskie), M. Cieplak,
"Chirality and protein folding",
J. Phys. Cond. Mat. (2005) 17, S1565.

Publication in the field of rheological properties of microscopic particles

13. M. Atakhorrami, J. I. Sułkowska, K. M Addas, G. Koenderink, J. X Tang, A. J. Levine, F. C. MacKintosh, C. F. Schmidt,
***"Correlated fluctuation of microparticles in viscoelastic solutions: quantitative measurement of materials properties by microrheology in the presence of optical traps''*,**
***Phys. Rev. E*, 73, 061501 (2006).**

Monographs, conference proceedings, and other publications

14. M. Cieplak, J. I. Sułkowska
"*Structure based models of biomolecules: stretching of proteins, dynamics of knots, hydrodinamic effects, and indentation of virus capsids*"
Springer, A. Koliński, Chapter 8, New York, pp. 179-208, DOI 10.1007/978-1-4419-6889-0 (2011).

- **Book chapter**

15. J. I. Sułkowska, P. Sułkowski, P. Szymczak, M. Cieplak,
"Stretching the knotted protein YibK and its unknotted constructs",
Proceedings of the conference on "Knots and soft-matter physics", Kyoto University, Japan, (2009).

16. M. Cieplak, Sz. Niewieczerzal, J. I. Sułkowska, P. Szymczak,
"Stretching of biomolecules in structure based models",
Longmans Orient, Bangalore, India (2009).

17. M. Cieplak, J. I. Sułkowska,
"Rozciąganie molekuł białek - porównanie ich własności mechanicznych",
Kosmos, 55, 4 (2006).

Description of achievements described in the above publications

My other achievements are presented in publications [1-13], in one book chapter [14], and reviews and conference proceedings listed above. These works are devoted to four areas of research: i) prediction of protein structures (publication [1]); ii) construction of coarse grained models and their application in the analysis of mechanical properties of proteins (publications [2-10]); iii) the use of coarse grained models in the analysis of thermodynamic processes in proteins (publications [11-12]); iv) viscoelastic properties of polymers (publication [13]). The results of these works are summarized below.

i) Protein structure prediction

In order to perform its biological function a protein must form a unique three-dimensional structure. Such structure in principle can be determined experimentally, however we still know many more sequences of amino acids that form proteins, rather than their spatial functional structures. In view of very large costs of experimental analysis of proteins, it is desirable to develop theoretical methods of bioinformatics to predict the spatial structure of proteins. Once homologous sequences or folds are known, such methods may predict spatial structures quite well, as confirmed by the results of international competitions The Critical Assessment of Protein Structure Prediction (CASP).

In [1] I constructed a new, efficient hybrid method, the DCA-fold, to predict the spatial structure of proteins. This method relies on a construction of a surprisingly effective technique to predict interactions between pairs of amino acids, based on the co-evolution of pairs of amino acid sequences in homologous proteins and the Direct Coupling Analysis method (DCA). Even though similar techniques, based on the co-evolution of amino acids, have been previously used, only the DCA enables to predict 50 to 300 non-local contacts with 70-80% accuracy for a variety of protein domains [1]. In [1] I constructed a simple hybrid method to predict protein structures, called DCA-fold, which integrates DCA contacts with an accurate knowledge of local information (e.g. the local secondary structure), as well as virtues of Go-like models. In addition I constructed statistical potentials (based on available 3D protein structures) to simulate long-range interactions. I predicted the local structure using statistical potentials for (psi, phi) angles and graphical methods of imaging of contact matrices. Using this model I predicted the structure of single-domain proteins, with an average number of 150 amino acids, with a resolution of 1-3 Å (RMSD from the native state).

It is worth noting that [1] is one of the first two papers in this field, and the DCA approach is now widely used and still developed. The paper [1] has been read more than 9000 times according to the PNAS database, and quoted more than 70 times since 2013.

ii) Molecular dynamics, stretching of proteins and mechanical properties of proteins

Folding and unfolding routes of proteins, while quite well characterized based on analysis of thermodynamical processes, for most proteins still have not been determined. The processes of spontaneous (thermally excited) folding and unfolding take place in timescales of the order from milliseconds to minutes. Unfortunately performing simulations of protein folding in the timescale of minutes is still beyond the capacity of modern computers. Nonetheless, mechanical stretching techniques provide an alternative approach to understand mechanisms of folding and unfolding. These techniques reduce time of achieving a given state by several orders of magnitude in computer simulations, and enable to conduct experiments for almost a single protein molecule.

The results of my research on the characterization of the free energy landscape of proteins by means of simulations of mechanical unfolding are presented in the series of publications [2-10]. In

those works I constructed an optimal coarse Go-like model to study mechanical manipulations of proteins, I used this model to characterize mechanical properties of 17134 proteins deposited in PDB, and identified mechanically most resistant proteins. In order to identify an optimal Go model I constructed a set of 54 different models, and compared their predictions (obtained in simulations of mechanical manipulations of several dozens of proteins) with experimental data. The results of these investigations are described in publications [6,8]. In the publication [10] I performed the first ever comprehensive, theoretical review of mechanical properties of 17134 proteins. Based on these results I selected 134 mechanically most resistant proteins, and characterized the shape of their mechanical clamps, i.e. geometric motifs responsible for high stability of protein structure. In [8,9] I improved the optimal model and used it to conduct the second review of mechanical properties, based on an expanded set of proteins. In addition in [8] I identified the most resistant protein out of those of known spatial structure, and in [3] I confirmed these theoretical results in experiment. Moreover in [8] I identified a new geometric motif (mechanical clamp) responsible for an exceptionally large mechanical resistance. I summarized the results of all these reviews in the database and a web server "Bio-molecule Stretching Database" (BSDB), which is available at ***http://jowisz.ifpan.edu.pl/BSDB/***. The BSDB database is now widely used by theoreticians and experimentalists to predict mechanical properties of proteins.

In [7] I constructed a hybrid model to analyze conformational changes in large proteins, whose structures in transition states are poorly accessible due to long simulation times, or even beyond the reach of the fastest computers when explicit solvent simulations are performed. The model that I constructed uses the transition state conformations determined in mechanical manipulations within the SBM model, and then it predicts consecutive unfolding steps of a protein based on fluctuation amplitudes computed within the Gaussian Network Model (GNM).

The publication [2] summarizes my studies of mechanical properties of proteins by means of optical tweezers. Amazing advances in the resolution of optical tweezers in recent years enabled measuring mechanical unfolding pathways of proteins almost at the level of thermal fluctuations, which are responsible for the breaking of hydrogen bonds in proteins. In [2] I identified differences between interpretation of experimental and theoretical (i.e. computer simulations) data. In this work I also developed a theoretical model for the analysis of experimental data obtained in the landmark optical tweezers experiment, conducted by the group of Prof. Rief, the author of the paper "*Full distance-resolved folding energy landscape of one single protein molecule*" (2010), 107, 5, 2013-2018.

In [10] I conducted stretching simulations of 20 knotted proteins in a coarse-grained model, and found that mechanical unfolding pathways of knotted proteins are significantly different than unfolding pathways of knotted homo-polymers. Moreover I showed that the process of tightening of a knot (upon stretching of a protein) proceeds in steps, which can be identified as Levy jumps. Analyzing the geometry of a protein I showed that upon stretching the knot's ends jump to well defined locations associated with sharp turns, and ultimately the knot tightens in the native area. The results of these studies have been later confirmed in theoretical studies within coarse grained models by Prof. Szymczak (University of Warsaw), in simulations in an explicit solvent model by Prof. Dziubiella (Humboldt-University of Berlin, Germany), as well as experimentally by Prof. Rief (Technische Universität Muenchen, Germany).

iii) The use of coarse grained models in the analysis of thermodynamic processes in proteins

In publications [11,12] I examined the impact of the use of coarse grained models on processes of thermal folding, unfolding and stretching of proteins, and compared it with known results obtained in all-atom simulations or experimental investigations. In work [11], based on comprehensive computer simulations of thermal unfolding of proteins, I defined the unfolding time

and showed its low-temperature divergence. I showed that as the temperature decreases, the median unfolding time grows faster than according to the Arrhenius law. Moreover, based on analysis of folding and unfolding events, I showed that below a characteristic temperature contacts break at separate time scales and unfolding proceeds approximately in a way reverse to folding. In publication [12] I showed that several simple criteria of folding to a native state are not sufficient in the case of a simple structure based model (where each aminoacid is described only in terms of Cα atoms). In particular criteria such as the number of native contacts or the root mean square deviation distance away from the native state are not sensitive to a wrong chirality. In [12] I also showed that including in the hamiltian of the Go-like model a term which favors native values of the local chirality significantly reduces the number of incorrectly folded local structures.

iv) The viscoelastic properties of polymers

In [13], in a series of experiments with one and two optical tweezers, I analyzed Brownian motions of microscopic particles in viscous or viscoelastic fluids. In this work I also developed and programmed a model for the analysis of auto- and cross-correlated fluctuations of particles in optical traps. Measurements with two optical tweezers are currently commonly used to study unfolding of DNA, RNA, and proteins.