1

Thermodynamics of Protein Structure Formation and Function

Dan W. Urry
Bioelastics Inc., Vestavia Hills, Alabama
USA

1. Introduction

1.1 The birth of thermodynamics with the development of the steam-powered heat engine

Thermodynamics was born of the need to improve efficiency of the steam-powered heat engine in order that flooded salt mines of England could become more productive. The *water-to-vapor phase transition* provides the physical property whereby the steam-powered heat engine functions. Heat flows into the engine at the 100°C of the phase transition to effect a dramatic volume expansion. *For the steam-powered heat engine, heating causes expansion to perform mechanical work*. Principal contributors to the initial development of thermodynamics were Nicolas Léonard Sadi Carnot (1824), French physicist and military engineer who died of cholera in 1832 at the age of 36 and William Thomson (Lord Kelvin), a physicist and engineer of the University of Glasgow, whose contribution was in the period of 1840 to 1855 (Smith, 1977).

Looking back at this remarkable development, Prigogine and Stengers (1984a) state, under the section heading of "Heat, the Rival of Gravitation" that "Out of all this common knowledge, nineteenth-century science concentrated on the single fact that combustion produces heat and that heat may lead to an increase in volume; as a result, combustion produces work. Fire leads, therefore, to a new kind of machine, the heat engine, the technological innovation on which industrial society was founded." Heating water at 100°C converts water to steam, a phase transition, to an increase in disorder (in entropy). Perhaps Lord Kelvin's statement of the Second Law of Thermodynamics is most relevant to our concerns, which is "It is impossible to convert heat completely into work in a cyclic process." Greater efficiencies in the conversion of heat into work become possible when heat is poured into a system at the temperature of a transition. Biology utilizes a unique and unfailing two-component phase transition of protein-in-water, and biology does so with a particularly empowering twist made possible by the accuracy and diversity of its protein sequences.

1.2 The aqueous protein-based heat engine of biology

The heat engine of biology comprises a two-component system of protein-in-water. Heating the fully hydrated (soluble) protein effects a phase separation of hydrophobic association (an association of oil-like side chains) that results in contraction. As depicted in Figure 1A, a model protein of the repeating pentamer sequence, $(\text{glycyl-valyl-glycyl-valyl-prolyl})_{251}$, in water (cross-linked by γ-irradiation to form a transparent elastic-contractile sheet) is swollen

below the temperature of the transition and contracts on heating to raise the temperature from below to above the that of the phase transition. As seen in Fig. 1B, on heating the strip becomes transiently opaque, while contracting to lift a weight in the performance of mechanical work. *For the protein-in-water heat engine of biology, heating causes contraction to perform mechanical work*.

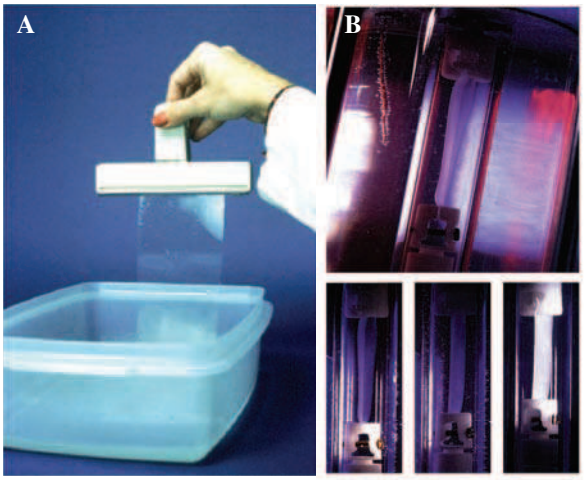

Fig. 1. An aqueous protein-based heat engine of biology, a water swollen sheet and a contracting strip of the cross-linked $(GVGVP)_{251}$, which is the basic elastic-contractile model protein of our study
A. Water-swollen transparent sheet below the temperature of the onset of the phase transition.
B. Upper: Aqueous chamber at a tilt containing a thermocouple and a strip, the heat engine, stretched by an attached weight. Lower: As the temperature is raised through that of the phase transition, the protein in water heat engine performs the work of lifting a weight by contraction. From Urry, 1995 with permission of Ann States Photography.

For warm-blooded animals, however, temperatures change very little. Importantly in these cases, the protein-in-water heat engine does not require heating to raise the temperature from below to above the temperature of the reversible phase separation of hydrophobic association in water to drive contraction. Instead contraction by hydrophobic association occurs by lowering the transition temperature from above to below body temperature, as attached biological functional groups are converted to their more hydrophobic states. The transition temperature is lowered by means of chemical or electrochemical energy inputs that convert a functional group from a more-polar to a more-hydrophobic state, such as occurs on charge neutralization or otherwise removal of charge. In mammals, when the temperature of the phase separation is lowered from above to below 37°C, contraction occurs as low entropy hydrophobic hydration becomes higher entropy bulk water (See section **9**: Summarizing Comments).
In your author's view, only when this increase in entropy (of pentagonal rings of hydrophobic hydration becoming less-ordered bulk water) is explicitly taken into consideration, can treatments of biological energy conversion involving changes in hydrophobic association in water be consistent with the Second Law of Thermodynamics.

That this performance of work, seen on charge neutralization, still represents an underlying protein-based heat engine is easily demonstrated. Here we note a family of model protein compositions that is considered in more detail below in Section **6**. At pH 7.5 in phosphate buffered saline, the glutamic acid (E, Glu) residue in **Model protein i**, (GVGVP GVGVP GEGVP GVGVP GVGVP GVGVP)$_{36}$GVGVP, is ionized as the carboxylate (-COO$^-$). This designed ECMP contracts when the temperature is raised from 55 to 70°C. For **Model protein i** lowering the pH to 3 forms the uncharged carboxyl (-COOH) and under this circumstance contraction occurs on raising the temperature from 15 to 30°C. Thus, at pH 7.5 **Model protein i** is a protein-in-water heat engine that contracts with a transition centered near 60°C, and at pH 3 **Model protein i** is a protein-in-water heat engine that contracts with a transition centered between 20 and 25°C. Thus, at pH 7.5 **Model protein i** *performs thermo-mechanical transduction* at elevated temperature, and at pH 3 **Model protein i** *performs thermo-mechanical transduction* below physiological temperature.

Also, **Model protein i**, at physiological temperature (37°C) and physiological pH, dissolves in water or occurs as a swollen cross-linked matrix. At 37°C, on lowering the pH to 3 the dissolved solution phase separates by hydrophobic association and the swollen cross-linked matrix contracts by hydrophobic association, with release of water, *to perform chemo-mechanical transduction*. Numerous functional groups of biology, attached to designed ECMP, drive contraction on conversion from their more polar state to their more hydrophobic state. Neutralization of charge results in formation of more hydrophobic hydration (See Figs. 10C and 12), with a negative $\delta\Delta H$ and a larger positive $\delta[-T\Delta S]$ (See Eqn. 4 of section **6.1.1** and associated discussion). This requires that the phase transition, where $\Delta H_t = T_t\Delta S_t$, occurs at a lower temperature. This ΔT_t-mechanism of energy conversion derives from input energies that shift the onset temperature, T_t, of phase transitions. The T_t-based Hydrophobicity Scale, of all amino acid residues in their different functional states (as applicable) and of additional functional groups, allows for the phenomenological design of ECMP capable of performing diverse free energy transductions (Urry, 2006a).

Experimental evaluations - 1) of the change in Gibbs free energy for hydrophobic association, $\mathbf{\Delta G_{HA}}$, to obtain a $\mathbf{\Delta G_{HA}}$-based Hydrophobicity Scale (Urry, 2004), 2) of an apolar-polar repulsive free energy of hydration, ΔG_{ap}, where charge disrupts hydrophobic hydration, and 3) of the mechanism of protein elasticity – allow insight into protein function, design of ECMP as transductional drug delivery/diseased cell targeting vehicles, and of many other medical and non-medical applications (Urry, 2006a; Urry et al., 2010).

1.3 Biology's inverse temperature transition, the rival of gravitation

Thus, for the biological world we note the Prigogine and Stengers (1984a) assertion that for the industrial world "Heat, the Rival of Gravitation" drives the phase transition of a more-ordered, condensed state of bulk water to the more-disordered, expanded gaseous state of steam to *achieve mechanical work by expansion*. And we extend it here to the biological world and argue that "Heat, the Rival of Gravitation" drives a phase transition to increased protein order by association of hydrophobic (oil-like) groups within and between protein chains to *achieve mechanical work by contraction*, (Urry, 1995; 1997; 2006a; Urry et al, 2010).

Central to understanding this phenomenon is that hydrophobic hydration is low entropy, structured water. Before the protein-in-water transition occurs, structured water arranges as pentagonal rings in association with hydrophobic groups (Stackelberg & Müller, 1951; 1954; Teeter, 1984), as may be seen in Fig. 2. During the phase transition of hydrophobic

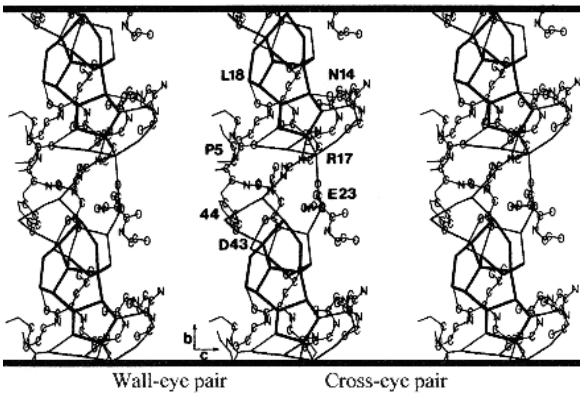

Fig. 2. Stereo views of residual pentagonal rings of hydrophobic hydration in association with hydrophobic moieties of L18 (leucine) and R17 (arginine) residues, after hydrophobic association of the small protein, crambin. From Teeter, 1984 with permission of M. M. Teeter.

association, the pentagonal rings of water of hydrophobic hydration become more-disordered as pentagonal rings of water become higher entropy bulk water (Urry et al., 1997). This decrease in order of water, i.e., increase in entropy, overwhelms in magnitude the increase in order on protein association, i.e., decrease in entropy, as hydrophobic groups of protein associate in the process of contraction (See section **6.1.3**). To emphasize this distinction, the ECMP-based phase transition to greater order of the model protein on raising the temperature is called an *inverse temperature transition*, (ITT). This is protein ordering on heating through the ITT of the ECMP, which ordering can be seen microscopically as the formation of twisted filaments that associate to form fibrils and fibers (Urry, 1992) and can even be seen with cyclic analogues of the model proteins as reversible crystallization on heating (Urry et al. 1978; Cook et al. 1980).

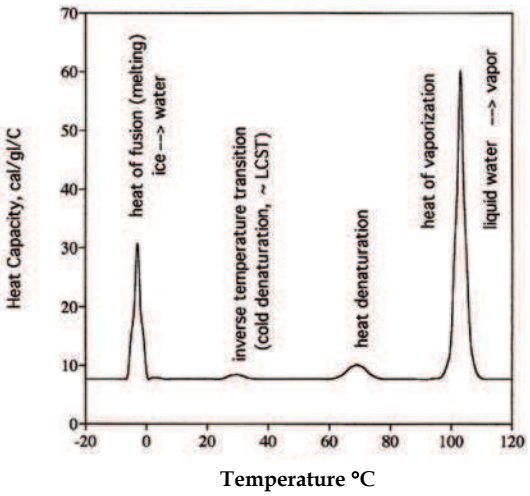

Fig. 3. Representation of endothermic phase transitions of $(GVGVP)_{251}$- in-water. From Figure 5.2 of Urry, 2006a.

Thus, without explicit consideration of water, which goes from being more-ordered to being less-ordered on raising the temperature from below to above the phase transition, the ITT of the protein-in-water heat engine of biology would seem to contradict the Second Law of Thermodynamics. But in fact, the heat driven increase in disorder (in entropy) as pentagonal rings of hydrophobic hydration become less-ordered bulk water is greater than the increase in order (decrease in entropy) as the model protein associates. Thus, in spite of the increase in order of the protein component, the ITT of ECMPs, is endothermic like those of the other transitions of water-to-vapor and ice-to-water, as water goes to a state of higher entropy of Fig. 3.

In summary, the water-to-vapor phase transition results in a dramatic increase in entropy of water and thereby enables the steam engine of the 19th Century Industrial Revolution to perform work by expansion. More profoundly, in your author's view, biology's inverse temperature transition results in a remarkable increase in entropy of water as pentagonal rings of hydrophobic hydration become higher entropy bulk water - whether driven by thermal energy input to raise the temperature through the phase transition or by chemical and other energy inputs that lower the temperature of the phase transition to hydrophobic association from above to below the operating temperature. This enables the diverse protein-based machines that sustain living organisms to perform work by contraction (Urry, 1995, 1997, 2006a; Urry et al., 2010).

1.4 Contrast between the arrow-of-time for the universe and the arrow-of-time for biology

Expressing his high esteem for the Second Law of Thermodynamics Eddington (1958) stated, "The law that entropy always increases – the second law of thermodynamics – holds, I think, the supreme position among the laws of Nature." With entropy measuring the increase in disorder, i.e., the increase in randomness, Eddington put forth the concept of "times arrow," (now commonly referred to as the arrow-of-time) using the argument, "Let us draw an arrow arbitrarily. If as we follow the arrow we find more and more of the random element in the state of the world, then the arrow is pointing towards the future; if the random element decreases the arrow points toward the past. That is the only distinction known to physics. I shall use the phrase 'times arrow' to express this one-way property of time which has no analogue in space. It is a singularly interesting property from a philosophical standpoint."

Considering the arrow-of-time, Toffler (1984), in the Forward to "Order Out of Chaos: Man's New Dialogue with Nature," (Prigogine & Stengers, 1984), addressed the dichotomy presented by biology with, "Imagine the problems introduced by Darwin and his followers! For evolution, far from pointing toward reduced organization and diversity, points in the opposite direction. Evolution proceeds from simple to complex, from 'lower' to 'higher' forms of Life, from undifferentiated to differentiated structures. And, from a human point of view, all is quite optimistic. The (biological) universe gets 'better' organized as it ages, continually advancing to a higher level as time sweeps by." The Toffler Forward set the stage for the Prigogine & Stengers thesis from the discipline of non-equilibrium thermodynamics, under which circumstances less-ordered systems may spontaneously give rise to complex more-ordered systems. Again quoting from Prigogine & Stengers, (1984b), "We can speak of a new coherence, of a mechanism of 'communication' among molecules. But this type of communication can arise only in far-from-equilibrium conditions. It is quite interesting that such communication seems to be the rule in the world of biology. It may in fact be taken as the very basis of the definition of a biological system."

Your author has previously argued (See the Epilogue of Urry, 2006a) that, while the energy required to produce the great macromolecules of biology is very large, the macromolecules themselves are not-so-far-from-equilibrium, due to discarding of 8 kcal/mol-residue with the addition of each residue. Yet repulsive free energies within complementary protein sequences can drive association between them. For further discussion of this issue see section **2**.

1.5 The components of this paper

Our perspective of the thermodynamics of protein structure formation and function unfolds below in seven parts: 1) Description of a key step in the biosynthesis of biomacromolecules, the nucleic acids and proteins, whereby biology achieves order out of chaos. The key step simply exemplifies an energy-fed reversal of biology's otherwise vaunted exception to the universal arrow-of-time. 2) Development of a model system of elastic-contractile model proteins (ECMPs) with which to establish the thermodynamics of hydration and of elasticity in protein function. 3) Phenomenological demonstration of a family of 15 pair-wise energy conversions achievable by designed ECMP capable of a thermally driven inverse temperature transition (ITT) to increased order by hydrophobic association. Thereby numerous inputs of intensive variables of the free energy - mechanical force, pressure, chemical potential, temperature, electrochemical potential, and electromagnetic radiation - act on different functional groups to change the temperature of the ITT. 4) Development of the thermodynamics of protein hydration ($\mathbf{\Delta G_{HA}}$ and ΔG_{ap}) and of elasticity (the internal energy, f_E, and entropy, f_S, components of force) as established by designed ECMP. 5) Noting how the Genetic Code (which is common to all characterized life on earth) facilitates protein-based machine evolution, new energy sources and improved machine efficiencies are, thereby, shown to be accessible at no increase in the energy required to produce new and/or more efficient protein machines. 6) The thermodynamics of protein hydration ($\mathbf{\Delta G_{HA}}$ and ΔG_{ap}) and of elasticity (f_E and f_S) are shown to be operative in biology's protein-based machines. 7) Application of the thermodynamics of Eyring's Absolute Rate Theory to the essential functions of trans-membrane transport processes of biology allows that the single image of the Gibbs free energy profile for ion passage from one side to the other of a cell membrane through a conduit of protein is sufficient to calculate trans-membrane ion currents as a function of ion activity and trans-membrane potential. This means of analysis, extrapolated to an array of essential biological trans-membrane transport processes, points to a future of a remarkable Eyring legacy, even to the trans-membrane transport processes of the energy factory of the living cell, the mitochondria of the animal kingdom and the chloroplasts of the plant kingdom.

2. How does biology reverse the universal arrow-of-time to achieve its order out of chaos?

In an early consideration relevant to biology's reversal of the universal arrow-of-time, Schrödinger (1944a) reasoned, "… we had to evade the tendency to disorder by 'inventing the molecule', in fact, an unusually large molecule which has to be a masterpiece of highly differentiated order.…" Almost a decade later Sanger (Sanger, 1952; Sanger & Thompson, 1953a; 1953b) demonstrated that proteins have specified sequences. *The means whereby biology achieves specified sequences for large chain molecules and the Genetic Code (See section 5) provide the solution as to how biology reverses the universal arrow-of-time, given sufficient energy*

supply. Anticipating construction of biological molecules different from anything as yet characterized by 1944, Schrödinger (1944b) further reasoned, "…from all that we have learnt about the structure of living matter, we must be prepared to find it working in a manner that cannot be reduced to the ordinary laws of physics." With remarkable foresight, he then went on to say, "… not on the grounds that there is any 'new force' or what not, directing the behaviour of the single atoms within a living organism, but because the construction is different from anything we have yet tested in the physical laboratory."

Indeed, a protein, in general, is in the words of Schrödinger (1944a) "an unusually large molecule" and always "a masterpiece of highly differentiated order." For a protein is a polymer, a polypeptide, in which each peptide unit may be formed of any one of 20 chemically and structurally diverse amino acid residues. So differentiated is the order that a 100 residue protein with the possibility of any one of twenty amino acid residues in each position gives the probability of a particular sequence as one in 10^{131}.

The key process in biology's reversal of the universal arrow-of-time resides within the synthesis of the magnificent macromolecules of biology, the nucleic acid and protein chain molecules of biology. These polymers exhibit precise sequences of subunits. The repeating units derive from four distinct nucleotides in each of the deoxyribonucleic acids (DNAs) and the ribonucleic acids (RNAs) and from 20 distinct amino acid residues of proteins. *Once these remarkably accurate sequences of diverse amino acids are obtained, three dimensional structure and function follow.* The primary structure, for example the accurate sequence of diverse amino acids of a protein, dictates protein folding and assembly, i.e., dictates three-dimensional structure (Anfinsen, 1973). Also, by the analysis reviewed here, the changes in structure that result in function, arise out of discrete energy inputs acting on biological functional groups attached to protein to bring about changes in hydrophobic association and often coupled with elastic deformation. Accordingly, an understanding, of how biology achieves order out of chaos and reverses the universal arrow-of-time, has as its basis an understanding of the thermodynamics whereby precise protein sequences are obtained, the Genetic Code, and the thermodynamics of protein function. In your author's view, central to understanding the energy conversions that constitute protein function are knowledge of the thermodynamics of hydrophobic hydration, elasticity, and Eyring Rate Theory.

2.1 A common key step whereby biology achieves order out of chaos in the biosynthesis for each of its great macromolecules – DNA, RNA, and protein

During construction of the nucleic acids and proteins of biology, the growing polymers are not-so-far-from-equilibrium. While protein and nucleic acid biosyntheses do require a very large amount of energy, the completed chain is *never*-very-far-from-equilibrium. The addition of each single amino acid residue for protein synthesis or of a triplet nucleotide codon of nucleic acid synthesis per amino acid, consumes ~24 kcal/mol of free energy. Discarding 24 kcal/mol to the environment, on adding each triplet codon to the growing nucleic acid and each amino acid residue to the growing protein chain, reproducibly produces accurate sequences. *A precise sequence dictates the three-dimensional structure of a protein in water for a given state of the functional groups of the sequence and of functional groups otherwise bound to the protein. And changes in state of the associated functional groups result in structural changes that give rise to function.*

In the biosynthesis of protein the activation of each amino acid (AA) and transfer to tRNA by aminoacyl-tRNA synthetase is given as follows: AA + ATP + tRNA = AA-tRNA + AMP +

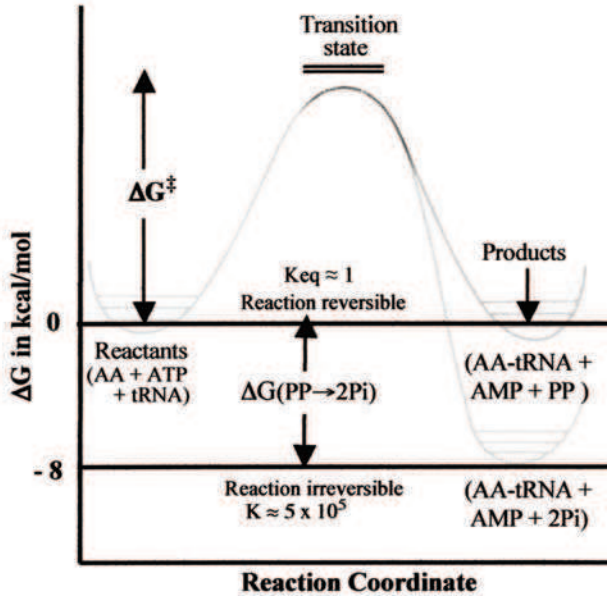

Fig. 4. Free energy profile for the reaction of amino acid (AA) plus ATP plus tRNA to produce the activated amino acid, i.e., AA-tRNA, ready for selective addition to the growing protein chain. The reaction may be seen in two steps: 1) The formation of AA-tRNA + AMP + PP, which is perfectly reversible with an equilibrium constant of one and the ratio of reactant to product of 1:1, 2) The enzymatic breakdown of pyrophosphate, PP $\rightarrow$ 2Pi + 8 kcal/mol, results in an irreversible overall reaction, i.e., K $\approx 5x10^5$. This very large cost of 800 kcal/mol-residue activation for production of a 100-residue-protein provides the free energy required for the peptide bond formation. There is yet another 1500 kcal-mol-(AA-tRNA) to bring the 100 AA-tRNA molecules out of disarray into alignment (see Eqns. 3b and 3c). Thus, some 2300 kcal/mol-residues added to take 100 amino acids (AA) out of chaos and to form a 100-residue protein of specified sequence.

PP(pyrophosphate), where AA stands for amino acid, ATP for adenosine triphosphate, tRNA for transfer-RNA, AA-tRNA for the activated amino acid as aminoacyl–tRNA energy-wise readied for addition to the growing protein chain, and PP for pyrophosphate. The equilibrium constant for this reaction required for attachment of each amino acid residue to tRNA is of the order of 1, i.e., K $\approx$ 1. The reactants and products occur at a ratio of approximately one. Due to the presence of an abundance of pyrophosphatase, catalytic breakdown of pyrophosphate immediately ensues, i.e., PP $\rightarrow$ 2Pi (inorganic phosphate) + 8kcal/mol. At each step of residue activation, a free energy of 8 kcal is released per mole of residue activated. As shown in **Figure 4**, this lowers the free energy of products by 8 kcal/mol. Based on this activation step alone, only one error would be made during the addition of some 500,000 residues. The free energy of pyrophosphate hydrolysis of 8 kcal/mol-residue-activated for addition to the growing chain immediately dissipates into the environment and is no longer associated with the process of chain growth. (For further discussion see Chapter 4 Likelihood of Life's Protein Machines: Extravagant in Construction Yet Efficient in Function of Urry, 2006a).

"Thus, (rather than employing far-from-equilibrium conditions) biology produces its macromolecules by means of an energetically extravagant, step-by-step, methodical march out of chaos" (See the Epilogue of Urry, 2006a).

2.1.1 Replication of DNA by G-C and A-T base pairings

Three steps lead to the biosynthesis of protein. These are: replication, wherein the strand of DNA that encodes protein sequence is duplicated for a daughter cell; transcription, the conversion of DNA into the equivalent sequence of RNA, and translation, the conversion of the ribonucleic acid sequence into the specified protein sequence. Beginning with replication of DNA, i.e.,

$$\text{parent DNA} \rightarrow \text{replication} \rightarrow \text{DNA of daughter cell} \tag{1}$$

An overall expression for DNA replication may be written as,

$$p\text{ATP} + q\text{GTP} + r\text{TTP} + s\text{CTP} = \text{DNA} + (p + q + r + s)\text{PP} \tag{1a}$$

where A (adenine), G (guanine), T (thymine) and C (cytosine) are the four bases, and the nucleotides - AMP (adenosine monophosphate), GMP (guanosine monophosphate), TMP (thymidine monophosphate), CMP (cytidine monophosphate) are the repeating units added one-by-one to form DNA. This applies to the synthesis of each strand of DNA to duplicate the DNA double helix. For biosynthesis of a 100-residue protein, the sum, $(p + q + r + s) = 300$.

A codon, which is a specific sequence of three bases, in general, encodes for one of the 20 amino acid residues, and there is a redundancy of codons for most amino acids. For example, there are four codons that encode for G (glycine, Gly) and a different four codons encode for V (valine, Val), and yet another set of four codons encode for P (proline, Pro), for A (alanine, Ala), and for L (leucine, Leu). On the other hand only one codon encodes for W (tryptophan, Trp) and six codons encode for R (arginine, Arg). The Genetic Code is a table that lists the codons that encode for each amino acid. As discussed in Section **5** below, the Genetic Code is arranged remarkably well for evolution of diverse and efficient protein-based machines that utilize modulation of inverse temperature transitions for function.

Again reaction (1a) occurs at near equilibrium for each nucleotide addition, but an abundant pyrophosphatase by way of reaction (1b) catalyzes the breakdown of pyrophosphate, PP, into 2 inorganic phosphates, 2Pi, and in the process releases 8 kcal/mol of energy to be dissipated into the environment, including heat that is no longer to be associated with the growing biomacromolecule.

$$(p + q + r + s)\text{PP} \rightarrow \text{pyrophosphatase} \rightarrow 2(p + q + r + s)\text{Pi} + (p + q + r + s) \times 8\ \text{kcal/mole} \tag{1b}$$

Thus, when encoding for a 100-residue protein, which requires a sequence of 300 nucleotides, there would be a free energy of (300 x 8) kcal/mol-residue released into the environment, that is, 2400 kcal/mol-300 base daughter cell DNA, which by transcription gives a 300 base strand of RNA, see Eqns. (2), as required for production of a 100-residue protein.

2.1.2 Transcription of DNA to produce RNA by G-C and A-U base pairings

The four bases of RNA are – adenine (A), guanine (G), uracil (U), and cytosine (C**)** – and the added nucleotide residues are – adenosine monophosphate (AMP), guanosine monophosphate (GMP), uridine monophosphate (UMP), and cytidine monophosphate

(CMP). The reaction constitutes transcribing a strand of deoxyribonucleic acid, DNA, into a strand of RNA. The statement of which may be given as Eqn. (2), i.e.,

$$\text{DNA} \rightarrow \text{transcription} \rightarrow \text{RNA} \tag{2}$$

The stoichiometry of the reaction may be given as,

$$\text{pATP} + \text{qGTP} + \text{rUTP} + \text{sCTP} + \text{DNA} = \text{DNA} + \text{RNA} +$$

$$2(\text{p} + \text{q} + \text{r} + \text{s})\text{Pi} + (\text{p} + \text{q} + \text{r} + \text{s}) \times 8 \text{ kcal/mol} \tag{2a}$$

Again, to encode for a 100-residue protein would mean (300 x 8) kcal/mol, or again 2400 kcal/mol being released to the surrounding solution.

2.1.3 Translation of RNA to produce protein

The translation of an RNA sequence into protein of $\eta = 100$, i.e.,

$$\text{RNA} \rightarrow \text{translation} \rightarrow \text{protein} \tag{3}$$

stated in terms of four reactions: a) The activation of an amino acid residue, AA_i, to its specific $tRNA_i$, discussed above, wherein AA_i, ATP, and $tRNA_i$ react to give AA_i-$tRNA_i$, AMP, and 2Pi with release of 8 kcal/mol-residue, i.e.,

$$\eta\ AA_i + \eta\ tRNA_i + \eta\ ATP \rightarrow \eta\ AA_i\text{-}tRNA_i + \eta\ AMP + 2\eta\ Pi + (\eta \times 8) \text{ kcal/mol-residue.} \tag{3a}$$

Eqn. (3a) represents a selectivity step where the correct amino acid is attached to its appropriate tRNA that contains the correct triplet codon for the amino acid being attached in an activated state. The amino acid selectivity process continues in the following reactions.

$$\eta\ AA_i\text{-}tRNA_i + \eta\ \text{ribosome(position 1)} + \eta\ GTP \rightarrow$$

$$\eta\ AA_i\text{-}tRNA_i\text{-ribosome(position 1)} + \eta GDP + \eta Pi + (\eta \times 7.5) \text{ kcal/mol-residue,} \tag{3b}$$

transfer to ribosome (position 2)

$$\eta\ AA_i\text{-}tRNA_i\text{-ribosome(position 1)} + \eta\ \text{ribosome(position 2)} + \eta GTP \rightarrow$$

$$\eta\ \text{ribosome(position 1)} + \eta\ AA_i\text{-}tRNA_i\text{-ribosome(position 2)} + \eta GDP$$

$$+ \eta Pi + (\eta \times 7.5) \text{ kcal/mol-residue} \tag{3c}$$

and finally the activated amino acid, AA_i-$tRNA_i$, bound at ribosome position 2, is added to the growing protein chain in its designated position in the sequence, i.e.,

$$\eta\ AA_i\text{-}tRNA_i\text{-ribosome(position 2)} \rightarrow \eta\ tRNA_i + \eta\ AA_i \text{ in protein} \tag{3d}$$

The cost in terms of Gibbs free energy to add a single amino acid to the growing protein chain is (8 + 2 x 7.5) kcal/mol-residue, and the cost of producing a 100-residue protein would be 2300 kcal/mol-100-residue protein.

As given above, the probability for a precise sequence of a 100-residue protein, with the possibility of one of 20 amino acid residues in each position, i.e., $(1/20)^{100} = 10^{-131}$. When the equilibrium constant is one, i.e., $K = 10^{-\Delta G/2.3RT} = 1$, ΔG is 0, and there is the probability of an

equal number of reactants and products. When the probability of a product is one chance in 10^{-131} for the occurrence of the product, one may write that $K = 10^{-\Delta G/2.3RT} = 10^{-131}$, or $\Delta G/2.3RT = 131$. Solving for the Gibbs free energy, $\Delta G = 131 \times 2.3RT = 186$ kcal/mole-100-residue protein. Calculated in this manner the efficiency of the synthesis of the 100-residue protein becomes 186/2300 = 0.08, i.e., an efficiency of the order of some 8%.

As will be noted below, protein-based motors can function at very high efficiencies. The F_1-ATPase (the F_1-motor of ATP synthase acting in reverse) has been calculated as approaching 100% (Kinosita et al., 2000). This has led to the exclamation that Life's protein machines are ***extravagant in construction yet efficient in function*** (See Chapter 4 of Urry, 2006a). (Some of the 1500 kcal/mol pays for a repulsive free energy between hydrophobic and charged groups.)

2.2 Precise primary structure, i.e., sequence, dictates three dimensional structure and function!

As argued above, a high price in terms of Gibbs free energy is paid in order to obtain polymers of precise sequence. Consequences of this severe price for precise sequence are the beautiful functional structures of biology. The more diverse the "side chains" of the repeating sequence, the more diverse are the functional capabilities. This is why the nucleic acids with but four similar repeating nucleotides each with the capacity of base pairing, i.e., A-T and G-C of poly(deoxyribonucleic acid) DNA and U-T and G-C of poly(ribonucleic acid), RNA, are suitable for sequence replication and transcription as considered above in terms of free energy required to produce precise sequences in Eqns. (1) and (2).

At the root of the structuring that becomes a living organism is the primary structure of DNA, the poly(deoxyribonucleic acid). DNA provides the sequence of bases that ultimately specify the precise sequence of protein. Protein sequence utilizes 20 structurally diverse residues that may be broadly classified as aromatic and aliphatic hydrophobic residues, as negatively and positively charged residues, and as neutral residues with non-ionizable polar functional moieties, capable, for example, of hydrogen-bonding. Overlapping with the latter two groups is cysteine with its –SH functional group that is commonly used in disulfide, -S-S-, cross-linking on formation of cystine.

Again, the probability of a precise sequence, with the possibility of one specific residue out of 20 residues in each position of even a relatively small 100-residue protein, becomes $(1/20)^{100} = 10^{-131}$, that is, one out 10^{131} sequences (See Chapter 4 of Urry, 2006a). This truly enormous number of possible sequences allows for an extraordinary number of protein three-dimensional structures with which to perform the diverse work (functions) required to sustain a cell.

2.2.1 Protein performs the work of constructing and maintaining the cell

The precise sequence of a protein, under physiological conditions, dictates the three-dimensional structure of the protein itself and whether it associates with like subunits to form an oligomeric protein comprised of symmetrically related subunits and/or with unlike subunits to form more complex protein structures. A remarkable example is ATP synthase of more than 20 subunits (10a, 2b, 3α, 3β, γ, ε). This rotary protein motor combines ADP (adenosine diphosphate) and Pi (inorganic phosphate) to make 32 of the 36 ATP (adenosine triphosphate) molecules on complete oxidation of a single molecule of glucose to 6 CO_2 plus 6 H_2O. Recall, ATP is the biological energy currency utilized to perform the work that sustains and propagates the living cell.

Assemblies of subunits, such as those of the three-fold rotary F_1-motor of ATP synthase, are dominated by hydrophobic inter-subunit interactions (Privalov, 1990) under the control of temperature and biological functional groups that can occur in two or more functional states. The more polar (e.g., charged) state, disrupts hydrophobic association and the more apolar (the more hydrophobic) favors hydrophobic association, each in a cooperative manner.

2.2.2 Familiar insights into the changes in hydrophobic associations that give rise to function

Insight begins with the familiar adage, "Oil and water don't mix!" Of course, they simply phase separate. But if oil-like and polar (e.g., water-like) groups are constrained to coexist along a polymer chain, they can't phase separate. Instead, the oil-like groups, dispersed along the polymer chain, self-associate by chain folding and by association with other chain segments, and, thereby, separate from water. But once the most favorable, the lowest free energy state, is obtained at a given temperature and pressure, only substantial changes in solvent or such as phosphorylation can change the state.

A related and more interesting adage becomes, "Oil and vinegar don't mix!" The solute of vinegar is principally acetic acid, which can exist in two states, i.e., CH_3-COOH ⇆ CH_3-COO^- + H^+, the very polar charged state, CH_3-COO^-, and the less-polar (more-hydrophobic) uncharged state, CH_3-COOH. Again as for oil and water, phase separation dominates the mixture of oil and vinegar. When hydrophobic and ionizable groups are forced by sequence to coexist as demonstrated with certain designed ECMP, it has been shown by means of substantial physical characterization of ECMP containing a glutamic acid (E, Glu) residue with the R-group of –CH_2-CH_2-COOH that the formation of the more polar state of –CH_2-CH_2-COO^- disrupts hydrophobic association (See for example Urry et al., 1997). It will be seen below, using the crystal structure of the closed conformation of the full-length KcsA potassium ion channel (Uysal et al., 2009) that the absence of carboxylate is seen associated with hydrophobic association that opens the channel, whereas the presence of carboxylate is seen associated with hydrophobic dissociation (Urry et al., 2010) and a closed channel. And the pH dependence of the conductance of the KcsA K^+-channel of Thompson et al. (2008) demonstrates conductance to turn off on the titration of glutamic acids to form charged glutamates.

Particularly, when the oil-like and charged groups are constrained to coexist by protein structure, they can be shown to reach out for hydration unperturbed by the other, that is, there is a competition for hydration between hydrophobic and charged residues (See for example Urry et al., 1997). This results in an *apolar-polar repulsive free energy of hydration*, ΔG_{ap}. (See Section **6.2.6** below and Urry, 1992; 1997).

2.2.3 Biological polymers of reproducible precise sequence add a new wrinkle to the *"laws of physics"*

Anticipating construction of biological molecules different from anything as yet characterized at the time, Schrödinger (1944b) further reasoned, "…from all that we have learnt about the structure of living matter, we must be prepared to find it (living matter) working in a manner that cannot be reduced to the ordinary laws of physics." With remarkable foresight, he then went on to say, "… not on the grounds that there is any 'new force' or what not, directing the behaviour of the single atoms within a living organism, but

because the construction is different from anything we have yet tested in the physical laboratory." Different constructions arise due to the capacity of biology to synthesize long proteins of precise sequence. This is because near physiological temperature the fundamental activation reaction, essentially independent of amino acid structure, has an equilibrium constant, K, of $10^{5.7}$. This translates into the order of one error in a half a million residue additions. Again, assuming that the twenty different residues possible at each position had an equal probability of being added, there would be 10^{131} different sequences possible for a 100-residue protein. This results in protein constructions that were simply inconceivable prior to the elucidation protein sequences and protein biosynthesis.

Again as Schrödinger (1944a) stated, "... living matter, while not eluding the 'laws of physics' as established up to date, is likely to involve 'other laws of physics' hitherto unknown, which, however, once they have been revealed, will form just as integral a part of this science as the former." As indicated above, the *new wrinkle to the "laws of physics"* derives from *an apolar-polar repulsive free energy of hydration,* ΔG_{ap}, that can be seen with the disparate side-chains (e.g., hydrophobic and charged) constrained to coexist along the precise sequence of which a protein chain is comprised.

It has been seen above that the reproducibly-achieved precise protein sequence (with an error as small as of one in one-half million residue additions) is achieved at an extraordinary cost in energy, and as such is consistent with the Second Law of Thermodynamics. It is not yet understood, however, just how the protein biosynthetic apparatus came into existence with which to achieve this protein construction so essential to the existence of life as we understand it.

2.3 Is the construction and maintenance of the biosynthetic apparatus for protein in accordance with the Second Law of Thermodynamics?

The biomacromolecular composition of the biosynthetic apparatus for production of protein requires RNA to specify protein sequence and protein catalysis to transcribe DNA into RNA, to produce tRNA, to attach amino acid (AA_i) to $tRNA_i$, i.e., to produce AA_i-$tRNA_i$, and to catalyze the steps in which the correct AA_i of an AA_i-$tRNA_i$ becomes attached to the correct position in the growing protein chain. The energy required for the latter, some 15 kcal/mol amino acid residue added of Eqns (3b) and (3c) in addition to precise protein sequence also pays for repulsive free energies that occur between disparate residues. DNA, RNA and protein chains of precise sequence are all simultaneously required in the first instance to achieve replication, transcription, translation to protein. How the initial biosynthetic apparatus came into existence is unknown. Once the ribosomal biosynthetic apparatus has been assembled with its accessory enzymes and nucleic acids all available, however, synthesis of protein does not contravene the Second Law of Thermodynamics.

3. A model protein system with which to establish thermodynamics of protein structure formation and function!

3.1 The composition of the basic model protein, $(GVGVP)_n$

Our model protein system, with which to establish thermodynamic elements of protein function, originates from the mammalian elastic protein, elastin, as a repeating pentapeptide sequence, $(GVGVP)_n$ with $n \leq 15$, depending on the species. A polypeptide chain may be represented as $[\text{-NH-CHR-CO-}]_n$ or as $[\text{-CO-NH-CHR-}]_n$ where the side chain (the R-group) of G (Gly, glycine) is the hydrogen atom, -H, the R-group of V (Val, valine) is $-CH(CH_3)_2$,

and the R-group of P (Pro, proline) is N_i–CH_2-CH_2-CH_2–C_i^{α}, i.e., three CH_2 groups spanning from the nitrogen atom, N, to the α-carbon of the same residue, i. Therefore, all side chains in $(GVGVP)_n$ are either the hydrophobic aliphatic hydrocarbons or the near neutral hydrogen atom and the only polar group is the recurring dipolar peptide moiety, -CO-NH-. Chemically and biologically synthesized $(GVGVP)_n$, with n ranging up to 200 or more, may be modified with sparse substitution of V by one or more functional groups, such as the carboxyls of glutamic and aspartic acids and the amino function of lysine (K, Lys), and additional biological functional groups such as redox functions, other prosthetic groups, phosphate, etc. Also, V residues may be replaced by the more hydrophobic F (Phe or phenylalanine) systematically to raise the hydrophobicity with the result of increased positive cooperativity giving an increased efficiency of energy conversion. These modified $(GVGVP)_n$ are called *designed* elastic-contractile model proteins (ECMP).

3.2 The molecular structure of the basic model protein, $(GVGVP)_n$

Figure 5A schematically represents the molecular structure of the basic model protein, $(GVGVP)_n$, as a series of VPGV β-turns with G spacers. Extending the spacer G residue to the adjacent V residue α-carbons, the VGV segment allows dynamic torsional oscillations of the intervening two peptide moieties. The damping of the amplitude of these peptide torsional oscillations gives rise to the librational entropy mechanism of protein elasticity (Urry et al., 1982d).

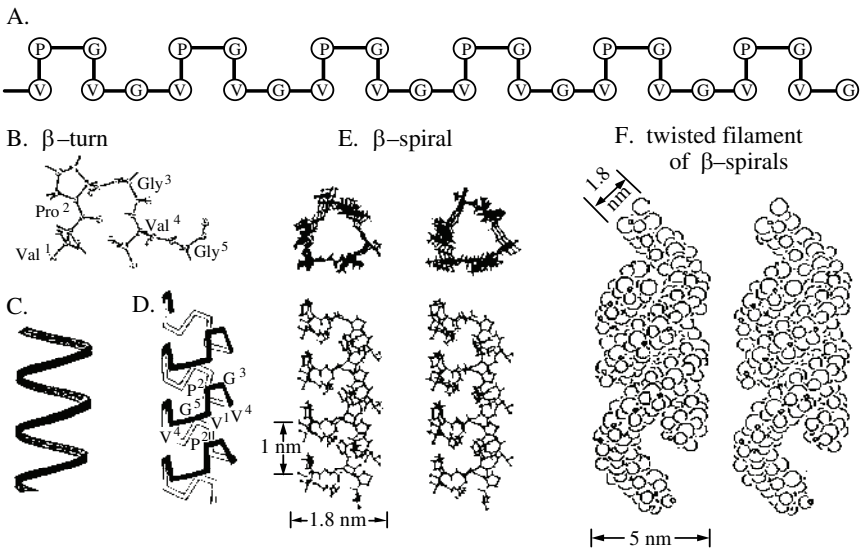

Fig. 5. Molecular structure of the elastic-contractile model protein, $(GVGVP)_n$. The structure is seen with repeating β-turns separated by dynamic suspended segments that wrap-up into associating β-spirals and exhibit simultaneous "near ideal" elasticity and phase transitional behavior from water to associate by hydrophobic interactions.
B. β-turn from the crystal structure of cyclo$(GVGVP)_3$ (Cook et al. 1980) which is the cyclic correlate of the linear β-spiral of D and E (Urry et al., 1981; Venkatachalam, et al. 1981; Venkatachalam and Urry, 1981). F. Adapted from Urry et al., 1982d.

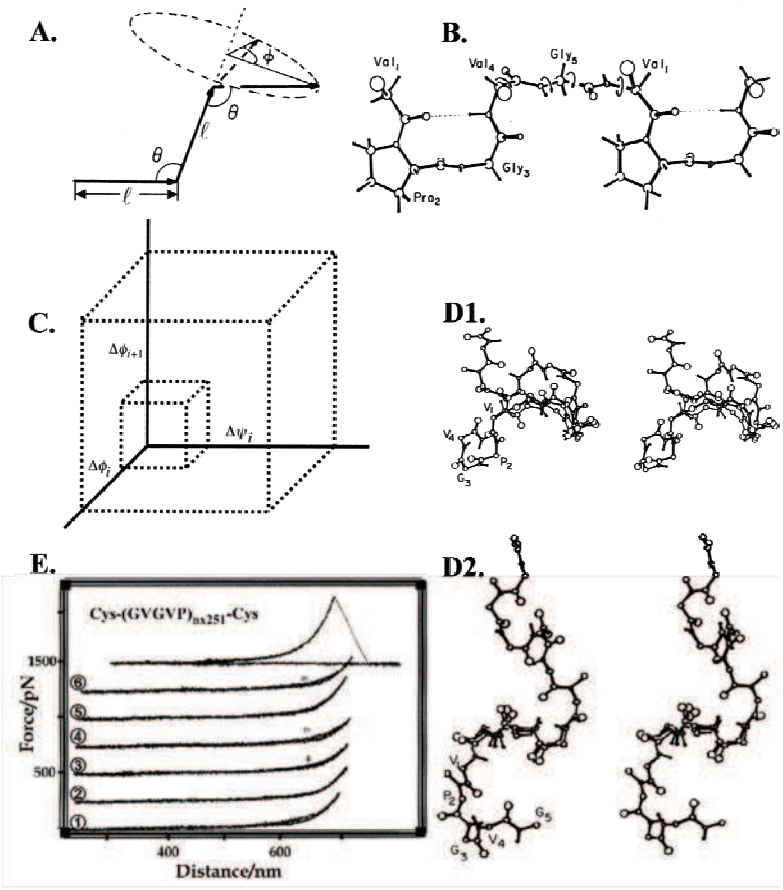

Fig. 6. Elements for understanding the nature of elasticity of the basic elastic-contractile model protein, $(GVGVP)_n$.

A. The structural elements for propagation of a rudimentary chain (Eyring, 1932). Bond length (ℓ), backbone bond angle (θ), torsion, or dihedral, angle (ϕ) due to rotation about bonds (Adapted from Urry, 1982).

B. Unrolled perspective of the β-spiral of $(GVGVP)_n$ showing suspended segments where peptide moieties are free to undergo large amplitude oscillations (peptide librations) involving the paired (ψ_4 & ϕ_5) and ψ_5 & ϕ_1 torsion angles. (From Urry, 1983)

C. Representation of entropy as a volume in configuration space, with axes plotting amplitude of torsion angle oscillations. As the volume increases due to larger torsion angle oscillations, a greater entropy can be calculated. (From Urry et al., 2010).

D1. One turn of the β-spiral with 2.7 β-turns per turn of spiral and an h of 3.5 Å showing the large torsion angle oscillations between the first and second β-turns. D2. On extension to an h of 8.0 Å, note damped oscillations. (From Urry & Venkatachalam, 1983).

E. Single-chain force-extension/relaxation curves, development of force during pulling in the z-direction of an AFM device with scans labelled from the bottom as Adapted from Urry et al., 2002.

The details of the β-turn are seen in **Figure 5B**, as obtained from the crystal structure of cyclo(GVGVP)$_3$ (Cook et al. 1980), which is the cyclic conformational correlate of the linear β-spiral conformation, as shown experimentally and computationally (Urry et al. 1981; Venkatachalam, et al. 1981; Venkatachalam and Urry, 1981). The linear β-spiral conformation is represented in increasing detail in **Figures 5C, D,** and **E** (Urry, 1990; 1991). Based on optical diffraction of negatively stained electron micrographs from incipient aggregates of dilute solutions (Volpin et al., 1976), three β-spirals of (GVGVP)$_n$ are thought to form twisted filaments as represented in **Figure 5F** (Urry et al., 1982d).

3.3 The unique properties of the basic model protein, (GVGVP)$_n$: "Near ideal" elasticity and phase transitional behavior

It is extraordinary that the basic model protein system, (GVGVP)$_n$ in water, simultaneously exhibits "near ideal" elasticity and thermally-elicited phase transitional behavior. To emphasize this unique and useful combination of properties, protein-based polymers based on (GVGVP)$_n$ have been given the descriptive name of elastic-contractile model proteins (ECMPs).

3.3.1 The "near ideal" elasticity of the basic model protein, (GVGVP)$_n$

An understanding of the "near ideal" elasticity of (GVGVP)$_n$ may be gained by discussing the component parts of Fig. 6, above. In particular, the curves of Fig. 6E utilized the basic atomic force microscope (AFM) (Hugel, 2003; Urry et al. 2002). Instead of imaging structures on a surface by rastering in the x- and y-dimensions, the cantilever tip moves in the z-direction with a long chain molecule spanning from the cantilever tip to the substrate surface to give a stress-strain curve that measures single-chain elasticity.

Ideal elasticity occurs when the plot of the force versus relaxation curve exactly overlays the force versus extension curve. Within the sensitivity (the noise level) of the measured stress-strain curves of Fig. 6E, the extension and relaxation traces of curves ② and ⑤ overlap. For curves ② and ⑤, therefore, the energy expended on extension is entirely recovered on relaxation, that is, these curves provide examples of ideal elasticity exhibited by extension and relaxation of a single-chain of Cys-(GVGVP)$_{502}$.-Cys. (Note: The Cys (cysteine, C) residues are present to achieve chemical (sulfhydryl) attachment to the cantilever tip of the AFM and for sulfhydryl attachment to the substrate surface.)

On the other hand curves ①, ③, ④, and ⑥ of Fig. 6E exhibit a higher noise level and in the original data separation is detectable as extension becomes greater than 600 nm. In these cases the extension curve is slightly higher than the relaxation curve, i.e., the cost in energy for extension is greater than the energy recovered on relaxation. Extension curves that occur at higher force levels than the relaxation curves are said to exhibit a hysteresis, which is an energy loss.

As seen in Fig. 4 of Urry et al. 2002 for Cys-(GVGIP)$_{260}$-Cys, the energy expended for extension is several times that recovered on relaxation. This is due in part to the greater hydrophobicity of (GVGIP) than of (GVGVP). The increase in the change in Gibbs free energy for hydrophobic association, ΔG_{HA}, results in a greater propensity for association with non-load bearing chain segments. A higher force on extension is required to disrupt these associations. The greater expenditure of energy to extend and disrupt these hydrophobic associations is not recovered on relaxation.

When the single-chain force-extension studies on Cys-(GVGIP)$_{260}$-Cys occur at very high dilution, however, essentially near ideal elasticity can be obtained. A slight hysteresis of

each of the curves, ①, ③, ④, and ⑥ of Fig. 6E, may be due to the chain folding back on itself, as time was allowed at low extension to increase the likelihood of backfolding (Hugel, 2003). As seen in Fig. 5E, the translation along the spiral axis for each complete turn is 1nm, and one complete turn requires three pentamers. Also, note in Fig. 6D1 that it is one turn of spiral, i.e., three pentamers, that is used in the calculation of the damping of torsion angle oscillations on extension by 130% from a value of 1 nm to 2.3 nm. Using the insight of Fig. 6C and the decrease in amplitude of torsion angle oscillation on extension, the change in entropy, ΔS, can be calculated by the equation, $\Delta S = R \ln[\Pi_i \Delta\phi_i^e \, \Delta\psi_i^e / \Pi_i \Delta\phi_i^r \, \Delta\psi_i^r]$.

The total elastomeric force, f_T plotted in Figure 6E, is the sum of an entropic component of force, f_S, and an internal energy component of force, f_E, i.e., $f_T = f_S + f_E$. The entropic component of elastic force is calculated as $f_S = -T(\partial S/\partial L)_{VT}$, where ∂S is calculated from the above expression for ΔS and ∂L derives from the 130% extension as used in Fig. 6D2. The sources of f_E derive from the reversible deformation of the angle, θ, and of the bond length, ℓ, both of Fig. 6A. (See section **6.2.10**, below.)

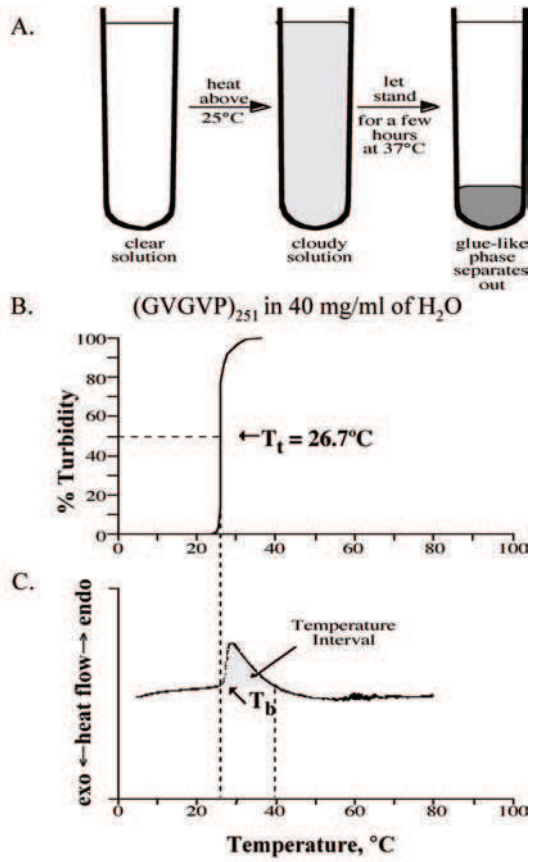

Fig. 7. Characterization of the $(GVGVP)_n$-in-water inverse temperature transition, using temperature dependence of turbidity and differential scanning calorimetry. From Urry, 1997.