Bulletin of Mathematical Biology Vol. 55, No. 6, pp. 1199–1210, 1993
Printed in Great Britain

0092-8240/93$6.00+0.00
Pergamon Press Ltd

COMPLEXITY OF PROTEIN FOLDING

■ AVIEZRI S. FRAENKEL*
Department of Mathematics,
University of Pennsylvania,
Philadelphia, PA 19104-6395, U.S.A.
(*E.mail: fraenkel@wisdom.weizmann.ac.il*)

It is believed that the native folded three-dimensional conformation of a protein is its lowest free energy state, or one of its lowest. It is shown here that both a two- and three-dimensional mathematical model describing the folding process as a free energy minimization problem is NP-hard. This means that the problem belongs to a large set of computational problems, assumed to be very hard ("conditionally intractable"). Some of the possible ramifications of this result are speculated upon.

1. Introduction. A protein is a sequence of amino acids, created as an essentially linear sequence from the well-understood genetic code inherent in a DNA chain, with semi-rigid bonds between pairs of amino acids which are adjacent in the linear sequence, and other, less rigid bonds between some other amino acids. The protein then folds into a complex three-dimensional *native conformation*, which determines its biological function. The folding mechanism is not known, but it is believed that the native folded conformation of a protein is its lowest free energy state (Anfinsen, 1973). Typically a protein consists of 1000–20,000 atoms and has a "diameter" of 35–100 Å (1 Å $= 10^{-8}$ cm).

One approach to model protein folding is to consider the protein to be a collection of hard impenetrable spheres (atoms) held together by elastic strings (covalent bonds). The atoms have electric charges that obey Coulomb's law (Levitt and Lifson, 1969). The electromagnetic force between any two atoms diminishes as d^{-2} and the potential energy as d^{-1}, where d is the distance between a pair of interacting atoms. The interaction is really between every atom of the protein and every atom in the universe! Yet within *ca* 1 sec the protein attains its final native three-dimensional conformation. Some chemical physicists simulate this system, where they will typically neglect forces between atoms at a distance >6 Å. Under this simplifying assumption simulation of 1 nsec (10^{-9} sec) of the protein folding process still takes some 150 hr on a modern main-frame computer (Levitt and Sharon, 1988). For more information about protein folding see e.g. Gierasch and King (1990).

* Permanent address: Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel.

In Section 2 we give some background on computational complexity and NP-completeness and in Section 3 we present a two-dimensional model of protein folding and prove it to be NP-complete. The result has been announced in Fraenkel (1990). We also indicate how the proof can be extended to a three-dimensional model. In the final Section 4 we speculate about possible ramifications of this and other NP-complete models of fragments of nature. Recently Unger and Moult (1993) have shown a three-dimensional protein folding model to be NP-complete.

2. Computational Complexity and NP-Completeness. The computational complexity of a problem is usually measured in terms of the number of "steps" or "time" to solve it, as a function of the problem's input size. A problem is called *tractable* if this function is polynomial; intractable otherwise. Thus, sorting n integers is tractable: it can be done in $O(n \log n)$ comparison steps. The following reasons motivate this convention:

(1) Normally, only tractable problems can be solved on a computer in reasonable time. Suppose that each of the problems π_1, π_2 and π_3 has input size n and that the best algorithms (="lower bounds") for solving them need n, n^2 and 2^n steps, respectively. If the rate of our machine is 10^6 steps/sec then for $n=60$, π_1 requires 0.00006 sec for execution, π_2 0.0036 sec and π_3 366 centuries!

(2) The input size of a problem that can be solved in a reasonable fixed time is of practical value only for tractable problems. Thus, for say 5 hr of uninterrupted computation π_1 can have input size 18×10^9; π_2 has size 13.4×10^4 and π_3 only 34. Moreover, any 10-fold gain in speed that technological advances may yield increases the size of a problem with an $O(n^k)$ algorithm that may be solved in a fixed time by the factor $10^{1/k}$, whereas the size of a problem with an $O(c^n)$ algorithm is increased only by an *additive* amount of $\log_c 10$ ($c>1$, a constant).

(3) The most simplistic approach to solving a problem is to explore its entire "search tree", i.e. searching through all possibilities. Except for trivial problems, this search constitutes an exponential algorithm. Thus, a problem whose best algorithm is exponential has often no essentially better algorithm than to search through all or most possibilities.

(In reality the world is not so simple; mostly in the pessimistic direction: there are problems which are polynomial and still appear to be intuitively intractable in two different senses! This is implied by the recent Robertson and Seymour (1988) theory in graph minors.)

Many problems can be shown to be tractable, simply by producing a polynomial algorithm for them. Some problems can be proved to be intractable. However, for the bulk of interesting problems both tractability and

intractability appear to be rather difficult to establish at present. For a large subset of them we can do the next best thing, which is to establish *completeness*, such as NP-completeness. For the moment we restrict attention to *decision problems*, i.e. problems for which the answer is YES or NO.

A decision problem π is NP-complete if:

(i) Given any solution for π, its validity can be verified in polynomial time (but there may not be a deterministic polynomial algorithm for finding a solution; we say that the problem has a "nondeterministic" algorithm). So the P of NP stands for Polynomial, N for Nondeterministic.

(ii) If π can be shown to be tractable then all NP-complete problems are tractable; if π can be shown to be intractable then all NP-complete problems are intractable.

Since the best known algorithm for any NP-complete problem is at present non-polynomial, all NP-complete problems are presently "practically intractable" or "conditionally intractable". For a thorough treatment of NP-completeness see Garey and Johnson (1979).

A common way to prove that a problem π is NP-complete consists of three phases:

(a) *NP-Membership.* Show (i) directly. This is usually, but not always, the easy part of the proof.

(b) *Construction.* Select an NP-complete problem π', consider an arbitrary generic instance x of π' and select a function f such that $f(x)$ is some instance of π and $f(x)$ is constructed in time which is polynomial in the size of x. This phase is normally called the polynomial *construction.*

(c) *YES-Equivalence.* Show that the answer to x is YES if and only if the answer to $f(x)$ is YES.

Phases (b) and (c) together are called a *reduction* of π' to π. Notation: $\pi' \propto \pi$.

The intuitive meaning of (a) and (b) should be clear: given an arbitrary instance x of π', it is transformed polynomially into a particular instance $f(x)$ of π (Fig. 1). If π is polynomial then, in particular, $f(x)$ can be decided in

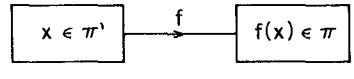

Figure 1. The intuitive meaning of a reduction.

polynomial time; the answer to $f(x)$ is YES if and only if the answer to x is YES. Thus, the polynomial algorithm for solving π and the polynomial f constitute a polynomial algorithm for solving π'. Equivalently, π' intractable implies π intractable.

Within the so-called *Turing machine* model (see e.g. Garey and Johnson,

1979) it is customary to denote by P the set of all tractable problems and by NP the set of all problems whose solutions can be verified in polynomial time. Then clearly $P \subseteq NP$. A major unsolved problem in theoretical computer science is whether $P = NP$ or not. It is customarily conjectured that $P \neq NP$. Any NP-complete problem belongs to the hardest problems in NP, in the sense that if $P \neq NP$ then the NP-complete problems are intractable.

Consider the problem:

Three-dimensional matching (3DM). Given three sets, X, Y, Z, with the same cardinality $|X| = |Y| = |Z| = q$, and a collection, $R \subseteq X \times Y \times Z$, does R contain a *matching*, i.e. a subset $M \subseteq R$ with $|M| = q$, such that for every two distinct triples (x_1, y_1, z_1), $(x_2, y_2, z_2) \in M$ we have $x_1 \neq x_2$, $y_1 \neq y_2$, $z_1 \neq z_2$?

Example. $X = \{x_1, x_2\}$, $Y = \{y_1, y_2\}$, $Z = \{z_1, z_2\}$, and $R = \{r_1 = (x_1, y_1, z_2)$, $r_2 = (x_1, y_2, z_1)$, $r_3 = (x_1, y_2, z_2)$, $r_4 = (x_2, y_2, z_1)\}$. Then $q = 2$, $k = 4$ and $M = \{r_1, r_4\}$.

Whereas two-dimensional matching (only two sets and R consists of ordered pairs of terms from these sets), also known as the *Marriage Problem*, is well-known to be tractable, the problem 3DM is NP-complete. See e.g. Garey and Johnson (1979).

3. NP-Completeness of Protein Folding. We consider the following two-dimensional model for protein folding.

Minimum free energy conformation of protein (MEP). Given a graph $G = (V, A)$ (with vertex set V and edge set A), $V \subset \mathbb{Z} \times \mathbb{Z}$ (i.e. V consists of lattice points in the plane), $A = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, a function $C: V \to \{-1, 0, 1\}$ (the charges), $K \in \mathbb{Z}^-$ (energy bound) and $L \in \mathbb{Z}^+$ (maximum distance). Is there a rearrangement V' of V with $V' \subset \mathbb{Z} \times \mathbb{Z}$ which preserves $d(\bar{u}, \bar{v})$ for all $(\bar{u}, \bar{v}) \in A_1$, such that $E \leqslant K$, where $E = \Sigma C(\bar{u})C(\bar{v})/d(\bar{u}, \bar{v})$, summed over all vertices $\bar{u} = (x_1, y_1)$, $\bar{v} = (x_2, y_2)$ with $(\bar{u}, \bar{v}) \notin A_1$ and $d \leqslant L$, where d is the discretized Euclidean distance $\lceil ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{1/2} \rceil$?

Our purpose is to show that the decision problem MEP is NP-complete. The corresponding optimization problem—where we ask for $\min(E)$ rather than only $E \leqslant K$—is then clearly not any easier than MEP. In general, if a decision problem π_1 is NP-complete then its corresponding optimization problem π_2 is said to be NP-*hard*. We note that NP-hard problems are not any easier to solve than the NP-complete problems they correspond to.

The NP-membership of MEP follows from the observation that, given any solution to MEP, we have only to check that $d(\bar{x}, \bar{y})$ is preserved for all $(\bar{x}, \bar{y}) \in A_1$ and that $E \leqslant K$, both of which can be done in time which is a polynomial in the input size of G.

We show 3DM$\propto$MEP. Let $X=\{x_1, \ldots, x_q\}$, $Y=\{y_1, \ldots, y_q\}$, $Z=\{z_1, \ldots, z_q\}$ and $R=\{r_1, \ldots, r_k\}\subseteq X\times Y\times Z$ be an instance of 3DM. We have to construct in polynomial time a graph $G=(V, E)$, $V\subset\mathbb{Z}^2$, K, $L\in\mathbb{Z}^+$ and a function $C: V\to\{-1, 0, 1\}$ such that R contains a matching $M\subseteq R$ if and only if there is a rearrangement V' of V with $V'\subset\mathbb{Z}^2$ which preserves $d(\bar{x}, \bar{y})$ for all $(\bar{x}, \bar{y})\in A_1$, such that $E\leqslant K$.

The basic building block for constructing G is a "square" subgraph G_1 (Fig. 2), consisting of four vertices on the corners of a unit lattice square, connected by four edges in A_1, forming a circuit. The charges on the vertices are indicated in the four corners, where $-$ denotes -1 and $+$ denotes 1.

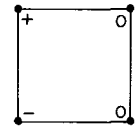

Figure 2. The basic building block G_1.

Using G_1 we construct the larger subgraphs of G: the *bed*, k *rods* and $k-q$ *caches*. The bed consists of a *bedstead* composed of two parallel horizontal bedstead rails, each of length $4(k+3q)+1$, the left part encasing the mattress of length $12q$, consisting of $3q+1$ vertical *cross bars*, partitioning the mattress into $3q$ *pillows*. The right part of the bed consists of the *bedpost* of length $4k+1$, containing k vertical *chests* of heights $10+3i$ $(1\leqslant i\leqslant k)$, measured from the upper bedstead rail. The k (vertical) rod *handles* have matching lengths. Each of the k (horizontal) rod *blocs* contains three copies of G_1, whose locations reflect $r_1, \ldots, r_k$: Order $X\cup Y\cup Z$ as follows:

$$X\cup Y\cup Z=\{x_1, \ldots, x_q, y_1, \ldots, y_q, z_1, \ldots, z_q\}.$$

If $r_i=(x_h, y_j, z_l)$ $(1\leqslant h, j, l\leqslant q)$, then the copies of G_1 on the bloc of rod R_i are at distances:

$$(4(k+3q-i+1-h)+1,\ 4(k+2q-i+1-j)+1,\ 4(k+q-i+1-l)+1)$$

from the rod handle of R_i $(1\leqslant i\leqslant k)$. Each rod handle contains two copies of G_1, in addition to one copy of G_1 (with four 0-charges) at the intersection of each rod handle with its rod bloc.

The $k-q$ caches are placed to the left of the left end of the bedstead rails, at a distance which is larger by 1 than the sum of the lengths of all the k rod blocs and handles from that left end. The charges are distributed as indicated in Fig. 3, which describes the global construction, which is complete by putting $L=1$ and:

$$K=-8(2k+3q).$$

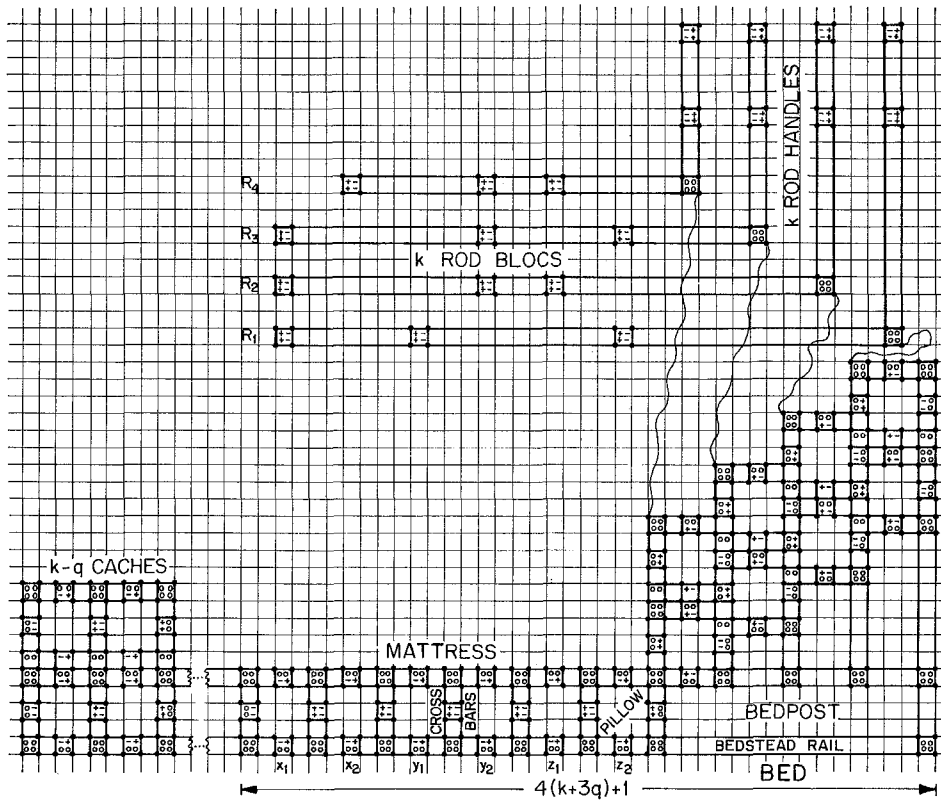

Figure 3. The global construction for the example at the end of Section 2.

Note that each of the three G_1-copies on the rod blocs has two $+1$ and two -1 charges, distributed as shown in Fig. 3. The pillows have a complementary charge distribution, so that each G_1-copy of a rod bloc, when embedded in the center of a pillow, will contribute a minimum of -8 to the energy. Analogously, the charge distribution on any rod handle is such that if it is embedded in a chest of the bedpost or in a cache it will also contribute -8 to the energy.

In Figs 2, 3 and 4 edges drawn horizontally or vertically belong to A_1, so they indicate that distances between their endpoints are preserved, whereas the other edges belong to A_2. Vertices connected by edges in A_2 and any pair of nonadjacent vertices (i.e. there is no edge between them) can move around to form V', subject to the constraints imposed by A_1-edges emanating from the pair to other vertices. Note that whereas V and V' are both embeddings in the planar lattice, the corresponding graphs are not necessarily planar. Thus, edges of A_1 may overlay. See e.g. Fig. 4, which describes the folded version of the

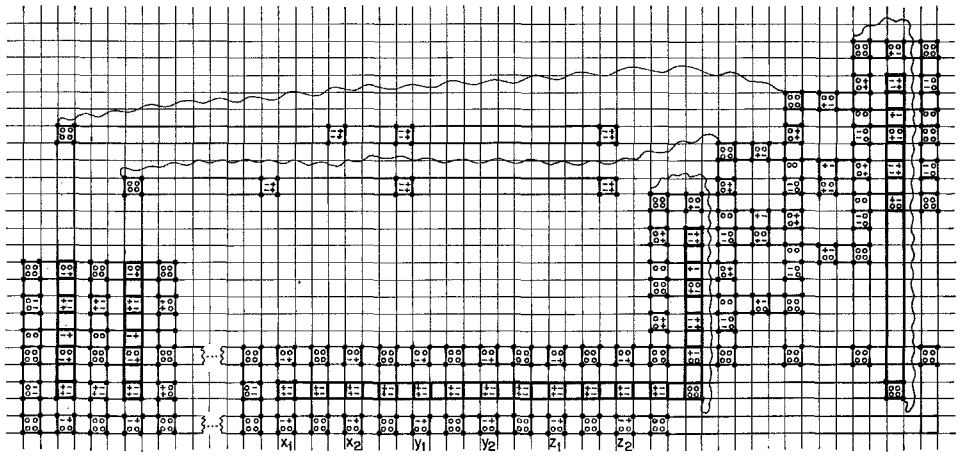

Figure 4. The folded version.

protein, where, e.g. edges of rod blocs overlay edges of cross bars and edges of rod handles overlay some edges of the caches.

Suppose that the given instance of 3DM has a matching $M=\{r_{i_1}, \ldots, r_{i_q}\} \subseteq R$. The locations of the G_1-copies on the bloc of rod R_i realize the three coordinates of r_i. Thus, the rods $R_{i_1}, \ldots, R_{i_q}$ can be embedded in the bed such that each pillow embeds precisely one G_1-copy and their handles are embedded precisely in the interiors of their matching chests. (In Fig. 3, R_1 and R_4 can be "lowered" and embedded in the bed in this way; the result is seen in Fig. 4.)

Each G_1-copy in a pillow or in a chest interior contributes -8 to the energy, so the contribution of the embedded $R_{i_1}, \ldots, R_{i_q}$ is $-40q$. The handles of the remaining $k-q$ rods can be embedded in the $k-q$ caches so as to contribute -16 each. These handles thus contribute $-16(k-q)$ to E. The total energy contribution is, therefore:

$$-40q-16(k-q)=-8(2k+3q)=K.$$

Now suppose that there is a rearrangement V' of V with $V' \subseteq \mathbb{Z}^2$ such that $E \leqslant K$. A rod whose bloc is embedded in the pillows and whose handle is embedded in its matching chest contributing -40 to E, such as R_1 in Fig. 4, is said to be *properly embedded* (in the bed). Each of the $k-q$ rod handles, tucked away in the $k-q$ caches (Fig. 4), contributes -16 to E. Some or all of these $k-q$ rod handles can be placed in some of $k-q$ chests, without their blocs necessarily occupying the pillows, also contributing -16 to E. A rod placed in a cache or in a chest such that the rod's handle contributes -16 to E is said to be *properly placed*.

If q rods are properly embedded and $k-q$ rods are properly placed then the

total free energy is precisely K, as we saw above. Moreover, the proper embedding of the rods $R_{i_1}, \ldots, R_{i_q}$ in the bed clearly implies that R has a matching $M=\{r_{i_1}, \ldots, r_{i_q}\}$. It suffices, therefore, to show that the free energy of any other rearrangement of V is larger than K (less than K in absolute value).

The proof is based on the observation that each vertex u of G_1 with a nonzero charge on a rod can contribute at least -2, and that this lower bound is attained if and only if u is at distance 1 from 2 opposite charges at right angles from u, which, if adjacent to u, must be adjacent via edges of A_2. Since a rod has 20 nonzero charges, the minimum free energy it can possibly contribute is thus -40, which is attained in a properly embedded rod. Since a rod handle has eight nonzero charges, the minimum free energy it can possibly contribute is -16, which is attained in a properly placed rod.

The only charged vertices at right angles which can interact with a charged vertex at distance 1 from them are in the pillows, chests and caches. Thus, the only rearrangement which can possibly attain E is to have the k rods interact with the pillows and the interiors of the chests or caches. This already excludes interaction of any two or more rods outside these interiors, although in a configuration of, say, three rods, a right angle with the above charge property yielding -2 to E can be made for some of the vertices on these three rods.

The geometry of the construction implies that the only way for a rod to contribute -40 to E is to be properly embedded. Therefore, it suffices to show that if q rods are properly embedded then the $k-q$ remaining rods, if properly placed, cannot be rearranged to contribute less than -16 each.

The point here is to observe that the bloc of a rod properly placed in a chest cannot interact with the rest of the chest to lower E. This follows from the fact that the charge distributions on the handles and on the blocs are complementary to one another. Thus, only the proper embedding of q rods and the proper placement of $k-q$ rod handles can contribute K to the free energy; and the proper embedding of the q rods in the $3q$ pillows implies that R has a matching M. This ends the proof.

Perhaps the graph of Fig. 3 does not appear to be very "linear" relative to the "folded" version of Fig. 4. However, the rods could have been drawn in Fig. 3 with all their blocs on one straight line, to the left of the bedstead rails. For elucidating the nature of the construction it was, however, more advantageous to draw the graph in a more compact form.

Our result holds also for a three-dimensional model of protein folding, defined the same way as MEP, except that $V, V' \subset \mathbb{Z}^3$, $\bar{u}=(x_1, y_1, z_1)$, $\bar{v}=(x_2, y_2, z_2)$ and $d=\lceil((x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2)^{1/2}\rceil$. The changes in the construction are straightforward: the rods are cubic with square cross section; the bed and caches are also three-dimensional, with two-dimensional projections shown in Fig. 3, but open on top so the rods can enter. The proof then goes through with $L=1$ and $K=-16(2k+3q)$.

4. Speculations About Ramifications. Whereas the first part of this paper was concerned with a *proof* of the NP-completeness of a decision problem, this section is nothing but a very brief speculation about nature. The immediate justification for indulging in this is nature's uncanny apparent ability to solve difficult problems. Here is one view:

(A) *Nature can solve* NP*-complete problems in polynomial time.* The following is but a small fraction of the supporting evidence.

Each amino acid in a protein can adopt, on average, eight different conformations (Privalov, 1979). A relatively small protein, consisting of 100 amino acids, can thus potentially assume 8^{100} conformations. Yet nature attains the native conformation in about 1 sec. (Note that the claim that nature assumes the global minimum free energy conformation in 1 sec is *not* equivalent to saying that it explores all the 8^{100} potential conformations in 1 sec!)

The double helices of DNA become knotted and linked in the course of biological processes such as replication. There are methods for realizing given knots and links on DNA chains (Wasserman and Cozzarelli, 1986). Moreover, the untying mechanisms follow topological transformations, so we can tell at the end of the process whether the unknot is a knot or not. The complexity of this question is a major unsolved problem in mathematics; see e.g. Welsh (1993).

In statistical mechanics many models have been studied to explain phase transitions (Baxter, 1982). Among them is the spin glass model: Given positive integers H, L and W, the three-dimensional grid graph $G=(V, A)$ whose vertices are the integer-coordinate points (x, y, z) with $1\leqslant x\leqslant H$, $1\leqslant y\leqslant L$, $1\leqslant z\leqslant W$, and whose edges connect each pair of vertices that are adjacent in one of the three directions, an integer interaction weight $J(a)\in\{-1, 0, 1\}$ for each edge $a\in A$ and an integer K. Is there an assignment of a spin $s(r)\in\{-1, 1\}$ to each vertex $v\in V$ such that the "ground state spin energy" E is $\leqslant K$, where $E=-\sum_{(u,v)\in A} J(u, v)s(u)s(v)$? This problem has been shown to be NP-complete by Barahona (1982). Special cases are polynomial; see Bieche *et al.* (1980), Barahona (1982) and Barahona *et al.* (1982). Nature usually manages to accomplish the phase transitions without a hitch and very fast.

Deutsch (1985, 1989) argues that "quantum computers" (computers based on quantum physics theory) can be constructed that can carry out in polynomial time computations which require exponential time. Bennett and Brassard (1989) proposed and implemented practical protocols—realizing an idea of Wiesner (1983)—and constructed a quantum counter for a certain public key cryptography application. See also Brassard (1988), Brassard and Crépeau (1990) and Bennett *et al.* (1992a,b).

In the cryptography quantum computer, polarized photons are used to

transmit digital information. Single photons (with high probability) are transmitted over a communication channel in one of four polarizations. There is no complexity gain in this device, yet it makes a dent in the Turing machine model: a Turing machine M can, by eavesdropping, learn the secrets being exchanged by two interactive Turing machines M_1 and M_2, without M_1 and M_2 knowing that M has learned their secrets. However, no machine can eavesdrop on two interactive quantum machines without being detected with high probability, assuming Heisenberg's uncertainty principle, which implies that eavesdropping to single photon transmission is tantamount to tampering with it.

Due to reasons of this type it was suggested in Fraenkel (1990) to try and reverse our usual scientific endeavor: in *addition* to modeling nature, studying the models, solving them or proving them to be NP-complete, etc., try to use devices of nature, such as proteins, quantum apparatus, DNA-chains, etc. as black boxes, to which we input instances of NP-complete problems and output solutions in reasonable time. One of the problems to be faced here is the design of efficient input/output interfaces to proteins and to other devices of nature. This is a problem even for the large DNA chains (containing some 10^9 nucleotides).

The opposite view is:

(B) *Nature functions within the Turing machine model.* How then can nature's apparent capability of solving NP-complete problems be explained without being forced to conclude $P=NP$? There are different answers to this question.

(i) Nature does not necessarily achieve *global* optimization. To this we may add, however, that often also a good stable local optimum is of interest to us.

(ii) NP-completeness is an asymptotic property, whereas the Universe seems to be finite. Moreover, it seems that once a protein is sufficiently large it is subdivided by nature into units of smaller size, say up to 200 amino acids per unit, which fold independently! See e.g. Privalov (1982) and Janin and Wodak (1983).

(iii) A problem π is polynomial if it is *universally* polynomial, i.e. if *all* its instances can be solved in polynomial time; it is NP-complete if *some* of its instances are NP-complete, although some of them may be solvable in polynomial time. In fact, NP-completeness reflects worst case behavior, but the average case behavior, under the assumption of some probability distribution, may be polynomial. Processes of nature are not necessarily universal. Thus, perhaps the natural selection of nature may help to preserve proteins with polynomial folding mechanisms and reject the others. Indeed, experiments show that some synthesized

proteins may fail to fold into a stable conformation. Similarly, perhaps nature creates only very few knot types in DNA helices.

(iv) The protein folding mechanism may be encoded in the protein's amino acid sequence, analogously to the genetic code of a DNA chain, but the code is still unknown. If this is the case then folding is not a search process and there is no issue of complexity.

I enjoyed conversations with Michael Levitt during 1990, when this work was done, and with Elisha Haas in 1992, when it was written up.

LITERATURE

Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science* **181**, 223–230.

Barahona, F. 1982. On the computational complexity of Ising spin glass models. *J. Phys. A: Math. Gen.* **15**, 3241–3253.

Barahona, F., R. Maynard, R. Rammal and J. P. Uhry. 1982. Morphology of ground states of two-dimensional frustration model. *J. Phys. A: Math. Gen.* **15**, 673–699.

Baxter, R. J. 1982. *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Bennett, C. H., F. Bessette, G. Brassard, L. Salvail and J. Smolin. 1992a. Experimental quantum cryptography. *J. Crypt.* **5**, 3–28.

Bennett, C. H. and G. Brassard. 1989. The dawn of a new era for quantum cryptography: the experimental prototype is working! *Assoc. comput. mach. SIGACT News* **20** (4), 78–82.

Bennett, C. H., G. Brassard, C. Crépeau and M.-H. Skubiszewska. 1992b. Practical quantum oblivious transfer. *Proc. Crypt.* **91**, 351–366.

Bennett, C. H., G. Brassard and N. D. Mermin. 1992. Quantum cryptography without Bell's theorem. *Phys. Rev. Lett.* **68**, 557–559.

Bieche, I., R. Maynard, R. Rammal and J. P. Uhry. 1980. On the ground states of the frustration model of a spin glass by a matching method of graph theory. *J. Phys. A: Math. Gen.* **13**, 2553–2567.

Brassard, G. 1988. *Modern Cryptology, Lecture Notes in Computer Science*, Vol. 325. New York: Springer-Verlag.

Brassard, G. and C. Crépeau. 1990. Quantum bit commitment and coin tossing protocols. *Proc. Crypt.* **90**, 49–61.

Deutsch, D. 1985. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. London A* **400**, 97–117.

Deutsch, D. 1989. Quantum computational networks. *Proc. R. Soc. London A* **425**, 73–90.

Fraenkel, A. S. 1990. Deexponentializing complex computational mathematical problems using physical or biological systems. Technical Report CS90-30. Department of Applied Mathematics and Computer Science, Weizmann Institute of Science.

Garey, M. R. and D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: Freeman.

Gierasch, L. M. and J. King (Eds). 1990. *Protein Folding: Deciphering the Second Half of the Genetic Code*. Washington, D.C.: American Association for the Advancement of Science.

Janin, J. and S. J. Wodak. 1983. Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. molec. Biol.* **42**, 21–78.

Levitt, M. and S. Lifson. 1969. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. molec. Biol.* **46**, 269–279.

Levitt, M. and R. Sharon. 1988. Accurate simulation of protein dynamics in solution. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7557–7561.

Privalov, P. L. 1979. Stability of proteins. *Adv. Protein Chem.* **33**, 167–241.

Privalov, P. L. 1982. Stability of proteins: proteins which do not present a single cooperative system. *Adv. Protein Chem.* **35**, 1–104.
Robertson, N. and P. D. Seymour. 1988. Graph minors XV, Wagner's conjecture, manuscript.
Unger, R. and J. Moult. 1993. Finding the lowest free energy conformation of a protein is a NP-complete problem: proof and implications. *Bull. math. Biol.* **55**, 1183–1198.
Wasserman, S. A. and N. R. Cozzarelli. 1986. Biochemical topology: applications to DNA recombination and replication. *Science* **232**, 951–960.
Welsh, D. J. A. 1993. The complexity of knots. *Ann. Disc. Math.* **55**, 159–171.
Wiesner, S. 1983. Conjugate coding. *Assoc. comput. mach. SIGACT News* **15** (1), 78–88 (manuscript written about 1970).

Received 26 March 1992
Revised 29 November 1992