From nuclear structure concepts to protein folding and non–conventional drug design

R.A. Broglia

Department of Physics, University of Milano, via Celoria 16, 20133 Milan, Italy
INFN, Milan Section, Milan, Italy
The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Abstract. Some of the paradigms emerging from the study of the phenomena of phase transitions in finite many–body systems, like e.g. the atomic nucleus, can be used at profit to solve the protein folding problem (how does a linear sequence of amino acids, immersed in the solvent, code for a unique, biological active, three dimensional native structure of the protein?), within the framework of simple (although not oversimplified) models. Also to design non–conventional drugs which do not create resistance (do not induce mutations in the virus or bacteria expressing the protein). The application of these concepts to the design of inhibitors of the HIV–1–PR, an enzyme which plays a central role in the life cycle of the HIV virus will be illustrated in terms of all–atom simulations and in vitro experimental results.

1. Introduction
Proteins are linear sequences of twenty different types of amino acids. They are produced by the ribosomes of cells following the blueprint of the genetic information. They fold into a unique, biologically active, three–dimensional (native) conformation in typical times ranging from 1μs to 1s. It is well established that in the folding process the protein undergoes an all–or–none transition from the denaturated (unfolded) state to the native, folded, state [1]. There also exists much evidence which testifies to the fact that this phase transition is completely determined by the linear amino acids sequence of the protein. Among the evidence one finds the fact that many proteins, especially small ones, denature reversibly, regaining their native structures spontaneously when returned to conditions that favour folding (see figure 1).

One of the great unsolved problems of science is the prediction of the three–dimensional structure of a protein from its amino acids sequence : the “protein folding problem” [1].

We shall show, in what follows, that concepts developed in the study of atomic nuclei to describe phase transitions of this finite many–body system can be used at profit to solve the protein folding problem, at least within the framework of simplified models. Because these models contain many of the features observed in real proteins, the insight obtained making use of these model concerning the mechanisms which are at the basis of the folding process, can be used at profit in the study of real proteins (see Sect. III, appl. to the HIV–1–PR).

1 Documentation of the plenary talk delivered at the Europhysics Conference on New Trends in Nuclear Physics Applications and Technologies, NPDC19, Pavia, Italy, September 5–9, 2005.
Figure 1. Denaturation of a typical protein as a function of concentration of denaturant, which changes the pH of the solvent. For a concentration of the denaturant of about 5M, the protein is denatured, while for concentrations \( \lesssim 1 \text{M} \) it is folded. It is also seen that the folding concentration, that is the denaturant concentration for which the probability for the protein to be in the folded state is equal to that of being in the unfolded state is \( \approx 3.5 \)M. Carrying measurements at this concentration, one obtains two peaks of equal area as a function of the parameter measuring the unfolded fraction. The similarity parameter \( q \) is the number of native contacts formed in a given situation, normalized with respect to the total number of native contacts.

Note that if one knows how a phase transition works, one also knows how to block it. But because proteins, to be functional (biologically active) have to be folded, to block their unfolded–folded phase transition is tantamount to blocking their function. If the target protein is one which plays an important role in the life cycle of a virus or bacteria, this means that one will be able to produce an efficient drug.

2. Phase transition
In a phase transition taking place in an infinite system, like for e.g. the normal–superconducting phase transition undergone by metals at the critical temperature \( T_c \), none of the particles involved play a special role (see figure 2(a)) while, as we shall see below, in the case of finite systems like the atomic nucleus phase transitions (e.g. normal–superfluid) are controlled by few, well defined, quantal states (see figure 2(b)). In other words, while in the description of phase transitions taking place in finite systems one does not need to make use of the singular functions used to describe infinite systems, one can describe the phase transition in terms of individual quantal states (see figure 2(c)) (cf. e.g. [2] and refs. therein).

Within this scenario it is useful to remember that the moment of inertia of rotational bands depend on the viscosity of the Fermi (non-newtonian) fluid inside the nucleus, and thus of the
value of the pairing gap. Independent particle motion displays a moment of inertia equal to the rigid moment of inertia, as the orbitals are solidly anchored to the rotating mean field. The nuclear Fermi superfluid (ground state configuration) displays, on the other hand, little viscosity, the associated moment of inertia being only between one-half to one-third the rigid moment of inertia.

Because the Coriolis force (analogue to the external magnetic field in metallic superconductors) acts differently on different Cooper pairs, strongly on those associated with “hot” (high-angular momenta) orbitals, weaker on “cold” (low-angular momenta) orbitals, the Coriolis-antipairing phenomena can provide information on the state dependent pairing gap $\Delta_j$.

Rotational bands associated with $\Delta = 0$ (normal, $N$), and $\Delta$ equal to the ground state value (superfluid, $S$) cross at a given value of the angular momentum (see figure 3).
Because of angular momentum conservation

\[ I_S\omega' = I_N\omega \, . \]  

Thus,

\[ \omega' = \frac{I_N}{I_S}\omega \approx 2\omega \, , \]  

implying a corresponding decrease (speeding) in the period \( T = \frac{2\pi}{\omega} \)

\[ T' = \frac{1}{2}T \, , \]

as shown in figure 4. Consequently, backbendings in the deexcitation of the nuclear system along the yrast line (see figure 5 as well as figure 2(c)) imply the extent to which the associated state dependent pairing gap associated with the band crossing contributes to the total pairing gap, and the role played by the corresponding Cooper pair in the condensed phase.

It has been found that, blocking the corresponding orbitals, no backbending is observed.

2.1. Denaturated-folded phase transition in proteins: hot amino acids and folding nucleus

The functional properties of proteins depend upon their three-dimensional structures. This native structure arises because particular sequences of amino acids (see figure 6) in polypeptide chains fold to generate, from linear chains, compact domains with specific three-dimensional structures (see figure 7).

As seen from figure 8, where an all atom representation of the HIV–1–PR, an enzyme central in the life cycle of the HIV virus is displayed, one is simply forced to simplify the complexity of the problem, in particular the force acting between amino acids.
Figure 4. Plot of the rotational period versus time associated with the yrast line of the nucleus $^{158}$Er.

Figure 5. Plots of the moment of inertia (top), spin (middle), and spin alignment (bottom) versus the rotational frequency for the yrast sequence in $^{158}$Er.

A possible simplification scheme is shown in figure 9 which essentially implies to represent the 20 different types of amino acids by beads all of the same size occupying the vertices of a cubic lattice.

Within this scenario, the protein folding problem is turned upside down into the quest for the sequences of amino acids which fold on short call into a selected native conformation (inverse folding problem [6]). This problem has, at least for small, single domain proteins, a simple
Proteins are built up by amino acids that are linked by peptide bonds into a polypeptide chain.

The amino acid sequence of a protein’s polypeptide chain is called its primary structure. Different regions of the sequence form local regular secondary structure, such as alpha (α) helices or beta (β) strands. The tertiary structure is formed by packing such structural elements into one or several compact globular units called domains. The final protein may contain several polypeptide chains arranged in a quaternary structure. By formation of such tertiary and quaternary structure amino acids far apart in the sequence are brought close together in three dimensions to form a functional region, an active site.

Solution: good folder sequences are characterized by a large gap $\delta = E_n - E_c$ (compared to the standard deviation $\sigma$ of the contact energies between the amino acids) between the energy $E_n$ of the sequence in the native conformation and the lowest energy $E_c$ of the conformations structurally dissimilar $^2$ to the native conformation [7-11], $E_c$ being a quantity which is solely determined by the composition of the protein, i.e. by the (experimental) relative presence of the different types of amino acids. In other words, good folders are associated with a normalized

$^2$ That is, conformations for which the value $q$ of the relative native contacts is $\leq 0.6$, $q = 1$ indicating the native structure (see also caption to figure 1).
gap $\xi = \delta/\sigma \gg 1$, quantity closely related to the z-score [12]. Furthermore, starting from a designed sequence which displays a large gap, all mutated sequences which preserve (to some extent) the gap fold into the native conformation [13]. Good folders are obtained by minimizing the energy of the chain in the native conformation with respect to amino acid sequence for fixed composition.

Within the lattice model, the configurational energy of a chain of $N$ monomers is given by

$$E = \frac{1}{2} \sum_{i,j}^{N} U_{m(i),m(j)} \Delta(|\vec{r}_i - \vec{r}_j|),$$

(4)

where $U_{m(i),m(j)}$ is the effective interaction potential between monomers $m(i)$ and $m(j)$, $\vec{r}_i$ and $\vec{r}_j$ denote their lattice positions and $\Delta(x)$ is the contact function. In Eq.(14) the pairwise interaction is different from zero when $i$ and $j$ occupy nearest–neighbour sites, i.e., $\Delta(a) = 1$ and $\Delta(na) = 0$ for $n \geq 2$, where $a$ indicates the step length of the lattice. In addition to these interactions, it is assumed that on–site repulsive forces prevent two amino acids to occupy the same site simultaneously, so that $\Delta(0) = \infty$ (excluded volume ansatz). The folding of
the chain is simulated by Monte Carlo (MC) methods. We shall consider throughout a 20–
letters representation of protein sequence where \( U \) is a \( 20 \times 20 \) matrix. A possible realization
of this matrix is given in ref. [14] (table VI), where it was derived from frequencies of contacts
between different amino acids in protein structures. The model we study here is a generic
heteropolymer model which has been shown to reproduce important generic features of protein
folding thermodynamics and kinetics, independent on the particular potential chosen [11,15].
This is achieved by using the same potential to design sequences and to simulate folding³.
However, in using such an approach, one should keep in mind that the labelling of amino acids
(spherical beads all of the same size and with no side chain) is generic too and may be no obvious
relation between those labels and labels for real amino acids.

We shall now discuss some of the results of a Monte Carlo simulation study of the dynamics
of a 36–monomers chain characterized by a polymer sequence, denoted \( S_{36} \) (cf. figure 10(b)),
designed by minimizing the energy in the target (native) conformation shown in figure 10(a).
In the units we are considering (\( RT_{\text{room}} = 0.6 \text{kcal/mol} \)), the energy of \( S_{36} \) is \( E_{\text{nat}} = -16.5 \),
while \( E_c = -14 \). Although this is not the sequence of lowest–energy, in particular the sequence
displayed in figure 10(c) has an energy in the native conformation of \( -17.13 \), \( S_{36} \) has a sufficiently
low energy and a large value of \( \xi (= 8.33) \) so that it can encode for a protein.

Monte Carlo simulations of folding performed on \( S_{36} \) at \( T = 0.20 \) (in our units) and using
a standard algorithm described extensively in the literature [15,16], in which, at each MC step
[17,18], a monomer is picked up at random and end and crankshaft moves as well as corner
flips are considered, indicate that this designed chain folds in a rather short “time” of \( 8 \cdot 10^6 \)
MC steps, and that at \( T = 0.28 \) the folding time is even shorter, \( 6.5 \cdot 10^5 \) MC steps. The
fractional population of the native state corresponding to these two temperatures is 91% and
10%, respectively, to be compared with a population of 0.5 and of \( 10^{-5} \) for the heteropolymer
folding temperatures of \( T = 0.26 \) (temperature at which the probability for folding as well as
for unfolding is 1/2) and \( T = 0.40 \) (temperature at which bonds break essentially as fast as they
are formed due to thermal fluctuations) respectively.

In figure 11 we display the distribution of the similarity parameter (relative native contacts
\( q \), i.e. \( q = 0 \) unfolded, \( q = 1 \) folded) associated with MC simulations at \( T = 0.26 \). The
distribution displays two peaks associated with the denaturated (small \( q \)) and the native (large \( q \))
conformations of the designed \( S_{36} \) protein. These results indicate that lattice models, although
being a quite simplified representation of proteins, are able to account for some important
features of these polypeptides, in particular the all–or–none (phase) transition between the
denaturated (unfolded) and native conformations experimentally observed (cf. figure 1).

2.2. Local Elementary Structures (LES)
In order to study the way in which each amino acid (site) contributes to the stability of the native
conformation of a designed protein, e.g., the \( S_{36} \) sequence, point mutations were introduced by
replacing a single amino acid by an amino acid of different type. There thus exist 19 such
possible substitutions for each of the 36 amino acids of \( S_{36} \). Studying the evolution of the
resulting sequences in conformational space through Monte Carlo simulations, one arrives to the
complete characterization of the 36 sites displayed in figure 10. It is found that 27 sites can be
considered “old” sites (light brown beads), 6 “warm” sites (yellow beads, \( \#3,5,11,14,16,28 \)) and
only three hot sites (red beads, \( \#6,27,30 \)). Thus, about 75% of the heteropolymer chain admits
single error transcription in the designed (correct) amino acid sequence yielding altered chains
which still fold to the native structure. In fact, introducing multiple mutations in the cold sites
(by swapping amino acids in order to conserve composition), the resulting sequences (\( \approx 10^{30} \),
³ This point is of crucial importance to generalize model studies to realistic situations, in that if one would like
to carry out a realistic calculation of a real protein, one would need to know (calculate) the interaction between
all the amino acids in a full all–atom calculation.
Figure 10. (a) The conformation of the 36-mer chosen as the native state in the design procedure. Each amino-acid residue is represented as a bead occupying a lattice site. The design tends to place the most strongly interacting amino acids in the interior of the protein where they can form most contacts. The strongest interactions are between groups $D$, $E$ and $K$ (cf. (b)), the last one being buried deep in the protein (amino acid in site 27). Amino acids occupying “hot” sites (sites 6, 27, 30) have been represented by red beads, those occupying “warm” sites (sites 3, 5, 11, 14, 16 and 28) by yellow and those occupying cold sites by light brown beads in this figure and by green beads in figure 12. The local elementary structures (LES) formed by the amino acid sequences $S_1^4 \equiv (3, 4, 5, 6)$, $S_2^4 \equiv (27, 28, 29, 30)$ and $S_3^4 \equiv (11, 12, 13, 14)$ and stabilized by the contacts 3–6, 27–30 and 11–14 (drawn by continuous lines) are explicitly shown by shadowed areas. The contacts between the LES are shown by dashed lines. (b) Designed amino acid sequence $S_{36}$. (c) Designed sequence $S_{36}'$ corresponding to $E_n = -17.13$.

Figure 11. The equilibrium distribution of the order parameter $q$ at the folding temperature $T = 0.26$. 

b SQKWLGERGATRIADGDLFVNGTYFSCKIMEVHPLA

c YPDLTKWAMEAGKIFSVDACLNGEGIRQVTLSN
Figure 12. Snapshots of the folding of the sequence $s'_{36}$ (see figure 10(c)), whose energy in the native conformation is $E_n = -17.13$ carried out at $T = 0.28$. Starting from a random conformation (a), the system forms after $\approx 10^2$ MC steps local elementary structures (LES) (b), involving three sets of four amino acids (3–6, 11–14, 17–30), whose stability is provided by the bonding indicated by dotted lines. When the LES come together to form the folding nucleus (FN) (indicated by dotted and dashed lines) after $7 \cdot 10^5$ MC steps (c), the system folds to the native conformation after only $10^3$ MC steps (d). The amino acids participating in the bonding of the LES (dotted lines) are among some of the most strongly interacting amino acids, which occupy, in the native conformation (d), “hot” and “warm” sites indicated by red and yellow beads, respectively. The monomers number 1 and 36 of the sequence $s'_{36}$ are indicated for each conformation (see figure 10(c)).

all preserving to some extent the gap $\delta = E_n - E_c$ still fold to the same native structure [19]. Mutations in the warm sites only delay the folding of the protein but not, as a rule, the ability the protein has to fold to the native structure or to a structures very similar to it. Mutations in the hot sites of the chain, which amount to $\approx 8\% - 10\%$ of all the amino acids lead to protein denaturation (misfolding), that is block the (unfolded) denatured→native ($D \rightarrow N$) phase transition (cf. figure 1), in a similar way in which odd nucleons in the $i_{13/2}(\nu)$ and $h_{11/2}(\pi)$ orbitals block the normal→superfluid pairing phase transition in rotating nuclei (like, e.g., in the case of the nucleus $^{158}$Er, cf. figure 5).

As stated above, mutations in the other sites lead to sequences which still fold to the native conformations and thus qualify as good folders. The resulting families of analogous proteins (proteins folding to the same native structure but not having a common ancestor) display in common essentially the few amino acids which occupy the hot sites. Examples of these families are well known experimentally, an example being provided by the immunoglobulin family [20].

The hot amino acids not only determine the stability of the protein but also the hierarchy of native contacts formation through which the protein, starting from an elongated phase reaches the native conformation (cf. figure 12): a) formation, almost instantaneously of few local elementary structures (LES, i.e. hidden intermediates corresponding to incipient $\alpha$–helices and $\beta$–sheets, the secondary structures of proteins) stabilized by the interaction between the hot amino acids, b) formation of the minimum set of native contacts which brings the system over the major free energy barrier of the whole folding process resulting from the docking of the LES (i.e., formation of the post–critical folding nucleus (FN)), c) relaxation of the remaining amino acids onto the native structure shortly after the formation of the FN giving rise to a unique
system with an energy below $E_c$ [11,21]. Summing up, the folding of proteins is controlled by the corresponding hot amino acids through the LES, ultimate building blocks of this molecular LEGO [22,23]. In other words, the simple, most important feature common to all designed sequences folding to the same native structure is the presence of few, highly conserved, strongly interacting, hot, amino acids which stabilize the LES and which, in the native conformation, are buried inside the folding nucleus of the protein.

With the help of the results discussed above, one has been able to develop a strategy which allows to predict the three–dimensional native conformation of a model protein from its amino acid sequence, that is to solve the protein folding problem (within lattice models) provided the contact energies acting among the amino acids are known [24]. This strategy, called the three step strategy (3SS), is the first bona fide solution of the protein folding problem, as the solution of the inverse folding problem requires the knowledge of the native conformation.

The generalization of the 3SS to the case of real proteins implies, in principle, the detailed knowledge of essentially only the interaction between the hot amino acids of the protein under study. Such an interaction can be obtained by calculating ab initio the corresponding contact energies between pairs of amino acids (cf. e.g. [25,26]). It is an open question whether three– and many–body interactions (i.e., simultaneous interaction of more than two amino acids) are important within this scenario (additivity condition). Paramount within this context is the so called hydrophobic interaction, that is, the “effective” interaction acting among non–polar (hydrophobic) amino acids, due to the presence of the solvent.

In any case, we shall not dwell any more on this (central) subject of the program of our group concerning the protein folding problem, but go directly to the subject of drug design, a subject which is providing important information also concerning the validity of the 3SS in general, and on the role LES play in the folding process in particular.

3. Drug design

LES elementary structures are also at the basis of a protocol for non–conventional drug design proposed by us [27]. Conventional drugs perform their activity either by activating or by inhibiting some target component of the cell. In particular, many inhibitory drugs bind to an enzyme and deplete its function by preventing the binding of the substrate, that is the molecule which has to be acted upon by the enzyme. This is done by either capping the active site of the enzyme (competitive inhibition) or, binding to some other part of the enzyme, by provoking structural changes which make the enzyme unfit to bind the substrate (allosteric inhibition). The two main features that inhibitory drugs must display are efficiency and specificity. In other words, it is not sufficient that the drug binds to the target protein and reduces efficiently its activity. It is also important that it does not interfere with other cellular processes, binding only to the protein it was designed for. These features are usually accomplished designing drugs which mimick the molecular properties of the natural substrate. In fact, the pair enzyme/substrate have undergone millions of years of evolution in order to display the required features. Consequently, the more similar the drug is to the substrate, the lower is the probability that it interferes with other cellular processes. Something that this kind of inhibitory drugs are not able to prevent is the development of resistance, a phenomenon which is typically related to viral protein targets. Under the selective pressure of the drug, the target is often able to either mutate the amino acids at the active site or at sites controlling its conformation in such a way that the activity of the enzyme is essentially retained, while the drug is no longer able to bind to it. In keeping with this result and with the central role played by LES in the folding process of proteins, we suggest the use of short peptides with the same sequence as LES (p–LES) as non–conventional drugs which interfere with the folding mechanism of the target protein, destabilizing
it and making it prone to proteolysis. These drugs are efficient, specific and do not suffer from the upraise of resistance.

In fact, the very reason why LES make single domain proteins fold fast confers p–LES the required features to act as effective drugs, that is, efficiency and specificity. They are efficient because they bind to the (complementary) LES as strongly as LES building the folding nucleus do. Since LES are responsible for the stability of the protein, their stabilization energy must be of the order of several times $kT$. These peptides are also as specific as LES are. In fact LES have evolved over millions of years so as to prevent the upraise of metastable states and to avoid aggregation, aside from securing that the protein to fold fast. In figure 13 we display results of model calculations which provide a quantitative assessment of the validity of these expectations.

The possibility of developing non–conventional drugs for actual situations is tantamount to being able to determine the LES for a given protein. This can be done either experimentally (e.g. through molecular engineering [1]), or extending the algorithm discussed in ref. [24] making

---

*Figure 13.* (a) Schematic representation of the way a conventional (standard) drug acts on the protein active site. A non–conventional drug shown in (b) displaying a sequence identical to that of a LES, will interfere in the folding of the protein, not allowing the FN to form (see figure 12(c), as well as figure 15). (c) Equilibrium distribution of the order parameter q (relative ratio of native contacts) associate at the folding temperature $T = 0.26$ for the designed sequence $S_{36}$ in the presence of $n = 0$ (see figure 11), $n = 1, 2, \ldots$ and 4 peptides of the type shown in (b).

---

$^4$ Process in which a protein undergoes cleavage by an enzyme which cuts a covalent bond (backbone bond) with accompanying addition of water, –H being added to one product of the cleavage and –OH to the other.
use of a realistic force field. The resulting peptides can be used either directly as drugs, or as templates to build mimetic molecules, which eventually do not display side effects connected with digestion or allergies. A feature which makes, in principle, these drugs quite promising as compared to conventional ones is to be found in the fact that the target protein cannot evolve through mutations to escape the drug, as happens in particular in the case of viral proteins in response to conventional drugs, because the mutation of residues in the LES would, anyway, lead to protein denaturation (see e.g. figure 17(c)).

An important example of drug-resistance is connected with AIDS. In this case, one of the main target proteins, the HIV-protease (see figure 14), a dimer formed out of two identical chains each containing 99 residues and folding according to the LES paradigm discussed above (cf. e.g. [28]), is able to mutate its active site so as to avoid the effects of drug action within a period of time of 6-8 months from the beginning of the therapy [29,30].

Making use of empirically determined potentials to describe the interaction among amino acids a full atom Molecular Dynamics (MD) calculation in explicit solvent was carried out starting from the native conformation of the protease, and followed it through 100 ps. Although this time is negligible compared to the folding time of the dimer (of the order of milliseconds), it is still sufficient to provide information concerning the interaction among the amino acids. In particular, it provides information concerning the 151 native contacts of each of the monomers. Although this is not an \textit{ab initio} interaction, one can use it within the framework of the 3SS to

Figure 14. Ribbon and side chain representation of the homodimer HIV–1–PR, an enzyme which play a central role in the life cycle of the HIV virus. This protein is a three–state homodimer [28], that is, a protein build out of two identical chains (each containing 99 amino acids) which reaches the native conformation starting from the unfolded (U) situation by first folding each of the monomer (2N) and then dimerizing them (N$_2$). Indicated in the upper right corner is the dissociation constant of the enzyme at two values of the pH.
determine the LES and the FN, aside from the hot sites\(^5\).

The results indicate the presence of two closed LES, each composed of \(\approx 8 - 10\) amino acids, and one open containing 4–5 amino acids, all containing some strongly interacting highly hydrophobic (non–polar) amino acids (figure 15). These LES are found to be in highly protected regions of the protein, and the few drug induced mutations which affect them are all conservative (i.e. highly hydrophobic amino acids are replaced by dito).

Based on these results we have chosen one of the (closed) LES as the most likely HIV–1–PR inhibitor (drug) (i.e. that containing amino acids 83–93). We call the corresponding peptide \(\text{p–LES}_8\) [31].

Simplified all atom simulations (Go–model with the value of the 151 native contact strengths taken from MD calculations) have been carried out allowing to evolve the HIV–1–PR in presence of three p–LES\(_8\) (see figure 16). The results indicate that this peptide displays strong inhibitory properties (see figure 17). Calculation making use of other (control) peptides containing 10–11 amino acids with sequence identical to segments of the monomers of the HIV–1–PR different from those associated with LES have also been carried out. Contrary to what is observed in the case of p–LES\(_8\), no sizable effect concerning the ability the protein has to fold has been observed [31].

Experimental assays of the effects of three different peptides have been carried out in vitro. The results (see figure 18) confirm the theoretical predictions [32]. The measured inhibitory constant \(K_i \approx 2.8\) \(\mu\)M competes well with that of conventional (active site) oriented drugs. The main difference of the p–LES\(_8\) non–conventional inhibitor is that it is unlikely it can lead to escape mutants, that is create resistance. In fact, the mutations which the virus should introduce in the HIV–1–PR to avoid the action of p–LES\(_8\) will involve hot amino acids (e.g.

\(^5\) Note that when dealing with real proteins one needs to consider a variety of the degrees of freedom. The results of model calculations help in extracting from the myriad of these degrees of freedom, those which are central for the question under discussion.

---

**Figure 15.** HIV–1–PR monomer in the native conformation, where the (closed) LES composed of amino acids 23–33 (green) and 83–93 (red), are displayed.
amino acid #33), essential for the folding of the protein (see figure 17(c)). The net result will be a strong destabilization of the protein. A modification of the p–LES$_8$ peptide which still is highly efficient ($K_i \approx 8 \mu$M) but qualitatively more soluble than the original wild–type p–LES$_8$ is now being tested in infected cells at the Institute of Infectious Diseases and Tropical Medicine of Sacco Hospital (Univ. of Milano) (see figure 19). At the same time a USA patent request has been deposited by the University of Milan on our behalf (see figure 20(a)).

Once the basic principle is understood there is, \textit{a priori}, no need to understand the detailed folding mechanism of a target protein to design the non–conventional inhibitor which will block its folding without allowing for escape mutants. In fact, chopping a protein into small peptides each containing $\approx 10$ amino acids and carrying out a modest number (of the order of 20–30) of experimental assays to assess the inhibitory abilities of each peptide, one can determine those with the best inhibitory properties. In fact, some of these peptides will display a sequence essentially identical to a LES and will thus inhibit the folding of the protein$^6$. An example of this strategy is shown in figure 21. This method which have been demonstrated to be successful in a number of situations, will shortly be employed to search for non–conventional inhibitors of the hemagglutinin (HA) and of the neuraminidase (NA), proteins essential in the life cycle of the influenza virus. Based on LES and hot orbitals, these inhibitors are likely not to find any important difference between the standard H1N1 virus or the potential pandemic–inducing H5N1 type, as the HA of all these variants depend for their folding on essentially the same hot and warm amino acids as well as LES. In fact, mutations which make the virus not recognizable by the immune system are epitope mutations (active site–like mutations).

Within the context of the present general method, the University of Milan has filed an European patent request on our behalf (see figure 20(b)). The claims of the invention have been

---

$^6$ This is equivalent to the situation in which a given nucleus displays a number of backbendings. Even if we do not know the detailed structure of the system, one can block one after the other the different single–particle levels lying close to the Fermi energy. Those levels which block (suppress) backbending, correspond to hot orbitals.
**Figure 17.** (a) Schematic representation of the result of molecular dynamic simulations of the evolution of the HIV–1–PR monomer in the presence of three p–S₈ peptides. (b) Native population Pₙ of the monomer as a function of p–LES (≡ p–S₈). (c) Same but for the mutated monomer. Designed in red are the intervals of amino acids belonging to the two LES (see figure 15).

**Figure 18.** Schematic representation of the optical set up used to measure the enzymatic activity of the HIV–1–PR as well as the inhibitor properties of the p–S₈ peptide.

judged original and patent rights will eventually be granted.
Figure 19. Scientific and development collaborations aimed at assessing the ability p–LES have to lead to non-conventional drugs.

* Institute of Infectious Diseases and Tropical Medicine, University of Milano
  M. Moroni
  G. Galli
  S. Rusconi
  A. Clivio (Biochemistry)

* Institute of Molecular Food Science, University of Milano
  E. Ragg (NMR spectroscopy; internalisation)

* CNR
  R. Longhi (Peptide synthesis)
  G. Ottolina (Enzyme assays)
  G. Colombo (all atom simulations)

* Bracco Imaging, s.p.a.
  Milano, Italy
  Princeton, New Jersey, USA

* University of Milano
  G. Sironi (Provost)
  R. Tiezzi (Patent Office)

**PATENTS:**

**INHIBITOR OF THE FOLDING OF THE HIV-1-PR AS ANTIVIRAL DRUG**

The present invention relates to a novel class of peptide compounds endowed with inhibitory activity towards HIV-1-pancreas and, more in particular, it relates to the said peptide and their use as antiviral drugs.

**Background**

As human immunodeficiency virus type-1 Protease (HIV-1-FR) is an essential enzyme in the viral life cycle, it is well known in the art that its inhibition can lead to a control of the acquired immune deficiency syndrome (AIDS). The main properties inhibitory drugs must display are efficiency and specificity.

**METHOD FOR THE IDENTIFICATION OF PROTEINS FOLDING INHIBITORS**

The present invention relates to a method for the identification of inhibitors of the folding and thus of the biological function(s) of proteins and, more in particular, peptide inhibitors of the folding of proteins which are highly selective and which do not create resistance.

**Background**

The fact that proteins play a primary physiological role is well known in the art. Many efforts have been taken to employ proteins as therapeutic agents, as catalysts and also as suitable materials possessing specific properties.

Many diseases stem from mutations in proteins that cause them to lose functionality.

Figure 20. Summary of the USA (a) and European (b) patent requests filed in July 2005 by the University of Milan on our behalf (Broglia, Tiana and Provasi (a), Broglia and Tiana (b)).
Figure 21. Relative population of the native conformation of the src–SH3 protein in the presence of three peptides containing 6 amino acids with a sequence identical to the segment starting with site x of the protein (containing 60 amino acids) and ending at site x+5 [33].
4. Conclusions

In figure 22 we collect some of the publications which are at the basis of the research which have brought us from the concepts developed in describing the phase transition of atomic nuclei, to solving the folding problem within simple models and to develop inhibitors of real proteins which do not create resistance.

- R. A. Brogila, G. Tiana, S. Pasquali, H. E. Roman, E. Vigezzi, *Folding and Aggregation of Designed Protein Chains*, Proc. Natl. Acad. Sci. USA, 95(1998) 12930

- G. Tiana and R. A. Brogila, *Folding and design of dimeric proteins*, Proteins Struct. Funct. Gen., 49(2002) 82

- R. A. Brogila, G. Tiana and R. Berera, *Resistance proof, folding-inhibitor drugs*, J. Chem. Phys., 118(2003) 4754

- R. A. Brogila and G. Tiana, *Reading the three-dimensional structure of a protein from its amino acid sequence*, Prot. Struct. Funct. Desig., 45(2001) 421

- R. A. Brogila, G. Tiana, L. Sutto, D. Provasi and F. Simona, *Design of HIV-1-PR inhibitors which do not create resistance: blocking the folding of single monomers*, Protein Sci., 14(2005) 10

- R. A. Brogila, D. Provasi, F. Vasile, G. Ottolina, R. Longhi and G. Tiana, *A folding inhibitor of the HIV-1 Protease*, Proteins (in press)

Figure 22.

One can hardly expect better evidence testifying to the fact that the study of the nuclear structure provides not only precious knowledge to understand the structure and working of our Universe, but also universal concepts which are at the basis of the remarkable properties displayed by soft matter, properties which are at the basis of life on Earth.

I want to acknowledge the important role G. Tiana, D. Provasi, L. Sutto, F. Simona and F. Vasile have played in the protein research discussed above. Within this context, a number of collaborations have been essential as shown in figure 19.

Particularly thanks are also due to S. Rusconi, G. Colombo, R. Longhi and G. Ottolina.
[1] A. Fersht (1999) *Structure and mechanism in protein science*, W. H. Freeman, New York (1999)

[2] D. M. Brink, R. A. Broglia, *Nuclear Superfluidity: pairing in finite systems*, Cambridge University Press, Cambridge (2005)

[3] C. Kittel, *Introduction to solid state physics*, J. Wiley and sons, New York (1996)

[4] J. Burdel *et al.*, Phys Rev. Lett. **48** (1982) 530

[5] Y. R. Shimizu *et al.*, Rev. Mod. Phys. **61** (1989) 131

[6] E. I. Shakhnovich, Phys. Rev. Lett. **72** (1994) 3907

[7] R. Goldstein, Z. Luthey-Schulten and P. Wolynes, Proc. Natl. Acad. Sci. USA **89** (1992) 4918–4922

[8] A. Sali, E. I. Shakhnovich and M. Karplus, J. Mol. Biol. **235** (1994) 1614–1636

[9] E. I. Shakhnovich and A. Gutin, Protein Eng **6** (1993) 793–800

[10] J. Bryngelson, J. N. Onuchic, N. D. Socci and P. Wolynes, *Proteins:Struct. Funct. and Gen.* **21** (1995) 167–195

[11] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7** (1997) 29–40

[12] J. Bowie, R. Luthey-Schulten and D. Eisenberg, Science **253** (1991) 164–169

[13] G. Tiana, R. A. Broglia H. E. Roman, E. Vigezzi and E. I. Shakhnovich, J. Chem. Phys. **108** (1998) 757–761

[14] S. Miyazawa and R. Jernigan, Macromolecules **18** (1985) 534

[15] E. I. Shakhnovich, Folding and Design **1** (1996) R50

[16] P. H. Verdier, J. Chem. Phys **59** (1973) 6119

[17] H. J. Hilhorst and J. M. Deustch, J. Chem. Phys **63** (1975) 5153

[18] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys. **21**, 1087 (1953)

[19] R. A. Broglia, G. Tiana, H. E. Roman, E. Vigezzi and E. I. Shakhnovich, Phys. Rev. Lett. **82** (1999) 4727

[20] T. E. Creighton, *Proteins*, W. H. Freeman and Co., Newy York (1994)

[21] R. A. Broglia and G. Tiana, J. Chem. Phys. **114** (2001) 7267

[22] G. Tiana and R. A. Broglia, J. Chem. Phys. **114** (2001) 2503

[23] R. A. Broglia, G. Tiana, S. Pasquali, H. E. Roman, E. Vigezzi, Proc. Natl. Acad. Sci. USA, **95** (1998) 12930

[24] R. A. Broglia and G. Tiana, Proteins **45** (2001) 421

[25] D. Provasi, Ph. D. Thesis (2002) University of Milano (unpublished)

[26] R. A. Broglia, G. Coló, G. Onida, H. E. Roman, Springer Verlag, Berlin, Heidelberg (2004)

[27] R. A. Broglia, G. Tiana and R. Berera, J. Chem. Phys. **118** (2003) 4754

[28] G. Tiana and R. A. Broglia, Proteins **49** (2002) 82

[29] A. Wodawer and J. Vondrasek, Annu. Rev. Biophy. Struct. **27** (1998) 249

[30] A. G. Tomaselli and R. L. Henrikson, Biochim. Biophys. Acta **1477**, 189 (2000)

[31] R. A. Broglia, G. Tiana, L. Sutto, D. Provasi and F. Simona, Protein Sci., **14** (2005) 10

[32] R. A. Broglia, D. Provasi, F. Vasile, G. Ottolina, R. Longhi and G. Tiana, Proteins (in press)

[33] R. A. Broglia, G. Tiana, D. Provasi, L. Sutto and V. Perelli, (to be published)