



Doctoral Thesis

On the frequency of folding in totally random libraries of de novo proteins and RNAs obtained by phage display

Author(s):

Chiarabelli, Cristiano

Publication Date:

2006

Permanent Link:

<https://doi.org/10.3929/ethz-a-005364485> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 16402

**On the Frequency of Folding in Totally Random Libraries
of *De Novo* Proteins and RNAs Obtained by Phage Display**

A dissertation submitted to the
Swiss Federal Institute of Technology Zurich

for the degree of
Doctor of Natural Sciences

presented by

Cristiano Chiarabelli

Dipl. Pharmaceutical Chem. and Techn. – University of Ferrara

Born October 7, 1972

Citizen of Ferrara, Italy

accepted on the recommendation of
Prof. Dr. Donald Hilvert, examiner
Prof. Dr. Pier Luigi Luisi, co-examiner
Prof. Dr. A. Dieter Schlüter, co-examiner

Zurich, 2006

1. ABSTRACT/SOMMARIO

ENGLISH

It is known that proteins and RNAs existing in nature are only a tiny fraction of the theoretically possible ones; an important question is how and why they have been selected during evolution. According to the contingency theory, extant bio-sequences may be the result of the simultaneous interplay of several concomitant causes. The straightforward consequence is that nature has not fully explored the whole space of sequences and thus, there may be an entire universe of proteins and RNAs whose properties have never been sampled by Nature.

The following work aims to investigate theoretically possible structures which have not been selected by evolution and are therefore not present on our Earth (“Never Born Proteins” - NBP and “Never Born RNAs - NBRNAs). In particular, I attempt to assess whether and to what extent such polymers might be folded. A library (about 10^9 clones) of totally random polypeptides, with a length of 50 aa residues, has been produced by phage display and, starting from the corresponding DNA, a library (about 10^8 sequences) of totally random RNAs has been obtained.

To probe the fold of the random peptides an enzymatic assay has been developed. The rationale behind this assay is that folded proteins are more resistant to proteolysis than unfolded ones. Thus, a constant DNA region (9 bp long) codifying for the tri-peptide thrombin substrate (proline-arginine-glycine, PRG) has been placed in the middle of the random DNA region codifying for each random peptide. The PRG-encoding DNA represents the only constant region in an otherwise completely random sequence. To set up the methodology, the PRG site was cloned in different positions along the polyproline helix of a model protein, Avian Pancreatic Polypeptide (APP), in order to assess whether and to what extent the APP overall fold avoids PRG cleavage by thrombin.

Results showed that proteolysis was prevented when the PRG site was buried in the APP secondary structure, whereas the PRG site was digested when placed in an unstructured region.

Therefore, this methodology proved to be able to distinguish folded proteins from random coiled ones on the basis of their proteolysis resistance.

79 phagemid clones, randomly selected from the initially large library, have been incubated with the enzyme overnight. Surprisingly, it turned out that over 20% of this population is thrombin resistant, which indicates some kind of protection, probably due to folding. Analysis of the primary sequences of these clones shows no significant similarity to extant proteins, which indicates they are indeed totally *de novo*. Two clones, chosen among those which proved to possess a stable fold have been expressed, isolated and purified and their structural properties have been characterized by spectroscopic methods and in particular by circular dichroism. This data shows a stable three-dimensional folding, which is temperature resistant and can be reversibly denatured by urea. The consequences of this discovery within a library of “Never Born Proteins” are discussed within the framework of molecular evolution.

On the other hand, to investigate the spontaneous folding of random RNAs the same DNA library has been engineered to perform *in vitro* transcription of correspondent random RNAs. To determine whether a stable structure is a common feature among a random RNAs library, a number of RNAs have been tested individually, using S1 nuclease, for the presence of double helices and indirectly any possible tertiary structure. In fact, folded RNAs have been expected to be more resistant to S1 nuclease than unfolded molecules, namely the latter should have been degraded faster than the former. In addition, the nuclease resistance test have been performed at different temperatures (S1 nuclease has been specifically selected for this reason), to study the temperature dependence of RNA folding stability.

Within this framework, RNA molecules have been tested individually at five different temperatures, in order to evaluate their ability to form internal double helices and, thus, to acquire secondary and tertiary structures. Results have been used to assess the temperature of the main unfolding transition of individual RNAs and to determine the ratio of folded RNAs at each temperature. Finally, experimental data have been compared with computer simulations of RNA folding, to support the robustness and reliability of the enzymatic assay.

ITALIANO

È noto che le proteine e gli RNA naturali conosciuti sono solo una piccola frazione di quelli teoricamente possibili; una domanda importante che ci si pone è come e perchè essi siano stati selezionati durante l'evoluzione. In accordo con la teoria della contingenza, le biosequenze esistenti possono essere il risultato dell'azione simultanea di diverse cause concomitanti. La diretta conseguenza di tale ragionamento è che la Natura non abbia esplorato l'intero spazio di sequenze possibili, così potrebbe esserci un intero universo di proteine ed RNA le cui proprietà non sono mai state prese in considerazione dalla Natura.

Il seguente lavoro ha avuto come obiettivo lo studio delle proprietà strutturali delle strutture teoricamente possibili, che non sono state selezionate dall'evoluzione e che non sono presenti sulla Terra ("Never Born Proteins" – NBP e "Never Born RNAs - NBRNAs). In particolare, ho tentato di valutare se e in che misura tali polimeri potessero assumere una o più strutture secondarie tali da garantire una struttura stabile. Una libreria (circa 10^9 cloni) di polipeptidi a sequenza completamente casuale, con una lunghezza di 50 residui amminoacidici, è stata prodotta attraverso "phage display" e, partendo dal corrispondente DNA, si è ottenuta una libreria (circa 10^8 sequenze) di RNA a sequenza casuale.

La struttura secondaria e terziaria dei peptidi a sequenza casuale è stata studiata tramite un saggio enzimatico che permette di distinguere i peptidi dotati di una struttura secondaria da quelli che non ne presentano alcuna. La metodologia si basa sull'osservazione che i peptidi dotati di struttura secondaria e/o terziaria risultano più resistenti all'idrolisi enzimatica di quelli non strutturati. Per sfruttare questa caratteristica dei peptidi strutturati, una regione costante di DNA lunga 9 nucleotidi codificante per il tri-peptide substrato della proteasi trombina (prolina-arginina-glicina - PRG) è stata inserita nel mezzo della libreria di DNA a sequenza casuale. La metodologia è stata messa a punto utilizzando la proteina modello APP (Avian Pancreatic Peptide). Tale proteina è stata modificata al fine di costruire diversi mutanti che esprimessero il tri-peptide PRG in diverse regioni dell'elica prolinica di APP. I saggi di proteolisi hanno mostrato come il sito sia protetto dall'attività della trombina solo quando ospitato in una regione dotata di struttura secondaria. Al contrario, il sito PRG viene prontamente tagliato quando localizzato in una regione non strutturata. Tali risultati hanno confermato la possibilità di utilizzare questa metodologia per distinguere peptidi con struttura da quelli non strutturati basandosi sulla resistenza alla proteolisi sito-specifica.

79 cloni fagemidici, selezionati in maniera casuale dall'iniziale vasta libreria prodotta, sono stati studiati tramite la metodologia precedentemente descritta. Sorprendentemente, i risultati mostrano che più del 20% di tale popolazione risulta resistente a trombina, suggerendo la presenza di domini secondari e/o terziari strutturati e stabili. L'analisi delle sequenze primarie di tali cloni non mostra similarità significative con le proteine conosciute e presenti in banche dati, tanto da poterle considerare totalmente *de novo*. Due dei cloni risultati resistenti alla proteolisi sono stati espressi, isolati, purificati e caratterizzati tramite metodi spettroscopici ed in particolare attraverso dicroismo circolare. I dati mostrano una struttura terziaria stabile, la quale è resistente a incrementi di temperatura e viene reversibilmente denaturata da urea.

Le conseguenze di tali scoperte, all'interno di una libreria di "Never Born Proteins" sono discusse in termini di evoluzione molecolare.

Per quanto riguarda invece lo studio dell'acquisizione spontanea da parte di RNA casuali di una struttura terziaria, la libreria di DNA è stata disegnata e sintetizzata clonando a monte della sequenza di DNA casuale un promotore per RNA polimerasi. Per determinare se una struttura terziaria stabile fosse una caratteristica comune in una libreria di RNA casuali, un certo numero di RNA sono stati testati individualmente utilizzando la nucleasi S1, in modo da determinare la presenza di doppie eliche e quindi indirettamente quella di possibili strutture terziarie. Infatti, ci si aspettava che RNA strutturati fossero più resistenti alla nucleasi S1 rispetto a quelli non strutturati, in pratica questi ultimi sarebbero dovuti essere degradati più velocemente rispetto ai primi. Inoltre, i test di resistenza alla nucleasi sono stati condotti a diverse temperature (la nucleasi S1 è stata scelta per tale ragione) per studiare la dipendenza della stabilità strutturale degli RNA rispetto alla temperatura. Le molecole di RNA sono state testate individualmente a cinque diverse temperature, in modo da valutare la loro abilità nel formare doppie eliche e di acquisire strutture secondarie e terziarie.

I risultati sono stati usati per determinare la temperatura di transizione da strutturato a non dei diversi RNA e per determinare il rapporto di RNA strutturato ad ogni temperatura. Infine, i dati sperimentali sono stati comparati con simulazioni al computer delle strutture degli RNA, per supportare la validità e solidità del metodo enzimatico utilizzato.