Physica A 451 (2016) 162–170

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Autoantibody recognition mechanisms of p53 epitopes

J.C. Phillips

Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, United States

HIGHLIGHTS

- The earliest stages of cancers generate biochemically detectable autoantibodies.
- A massive clinical study showed that p53 epitopes are the most effective tool.
- Bioinformatic scaling identifies these epitopes as exposed beta strands.
- Refined p53 epitopes are shorter and can be obtained from other species.
- Liver cancer epitopes can be studied clinically with 1000 times fewer patients.

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form 30 December 2015
Available online 19 January 2016

Keywords:
Self-organized criticality
Scaling
Autoantibody
P53
Epitope
Beta strand

ABSTRACT

There is an urgent need for economical blood based, noninvasive molecular biomarkers to assist in the detection and diagnosis of cancers in a cost-effective manner at an early stage, when curative interventions are still possible. Serum autoantibodies are attractive biomarkers for early cancer detection, but their development has been hindered by the punctuated genetic nature of the ten million known cancer mutations. A landmark study of 50,000 patients (Pedersen et al., 2013) showed that a few p53 15-mer epitopes are much more sensitive colon cancer biomarkers than p53, which in turn is a more sensitive cancer biomarker than any other protein. The function of p53 as a nearly universal "tumor suppressor" is well established, because of its strong immunogenicity in terms of not only antibody recruitment, but also stimulation of autoantibodies. Here we examine dimensionally compressed bioinformatic fractal scaling analysis for identifying the few sensitive epitopes from the p53 amino acid sequence, and show how it could be used for early cancer detection (ECD). We trim 15-mers to 7-mers, and identify specific 7-mers from other species that could be more sensitive to aggressive human cancers, such as liver cancer. Our results could provide a roadmap for ECD.

1. Introduction

In general protein–protein binding occurs at "hot spots" which are usually enriched in tryptophan, tyrosine and arginine, and hydrophobic occlusion of solvent is found to be a necessary condition for strong binding [1,2]. Interest in specific molecular biomarkers for early cancer detection is growing because of evidence that suggests that autoantibodies stimulated by cancer cells may share specific paratopes that selectively bind to p53 epitopes [3–7]. The superior selectivity of p53 epitopes for autoantibody paratopes [8] suggests that more specific mechanisms may be involved, such as interactions with paratope aromatic side chains and hydrophilic residues [2,9]. Among all proteins p53 is much more hydrophilic than average, and it is also elastically much softer, with about half its structure in its center dominated by β strands [10,11], while the remainder (especially the N-terminal quarter) is disordered. Because we are focused only on p53 and a few of its

E-mail address: jcphillips8@comcast.net.

http://dx.doi.org/10.1016/j.physa.2016.01.021
0378-4371/

epitopes, we are able to avoid the enormous complexity of millions of cancer genomic mutations [12]. Multiple signaling pathways, and/or multiple driver gene mutations, may be reflected simply in different p53 epitopic response patterns. This is an extreme example of what mathematicians call dimensional compression.

For the reader's convenience we summarize here the main features of our recently developed bioinformatic scaling method. The entire approach is post-Newtonian, and initially it may appear counter-intuitive. The central starting point is the implicit information contained in amino acid sequences in NCBI and their related structures in the PDB. Quite generally, one suspects that buried in these data are several key features of Nature's solutions for specific proteins to Levinthal's paradox [13]. Newtonian all-atom simulation models that involve motions of thousands of atoms are not only extremely cpu intensive (typically months to reach a few results), but they do not address the evolutionary nature of protein sequence–structure–function relations, or provide tools for extracting functional information hidden in NCBI and PDB.

The first break in addressing this problem was the landmark discovery of a uniquely effective hydropathicity scale [14]. Competing effects of hydrophobic and hydrophilic segments of a given protein have long been known to be the primary driving force behind the folding of protein chains into protein globules. There are secondary effects associated with longitudinal hydrogen bonding (α helices) and transverse hydrogen bonding (β strands), and even weaker charge effects, but in most proteins the dominant physico-chemical factor in a property such as aggregation [15] is hydropathic interactions. This new scale focuses on the hydrogen bonds in the interface between the water film and the protein. It is based on the solvent-accessible areas of amino acid van der Waals surfaces partitioned in overlapping regions according to the Voronoi rules. It has fundamental value because it describes globular protein sequence–structure–function relations in the context of a thermodynamic model of structural transitions dominated by the rearrangement energies of the interfacial hydrogen bond network.

The second step is unexpected, but the justification of it has become apparent from studying many examples. The first popular application of hydropathicity scales was to the study all-α heptad transmembrane opsins, whose seven internal transmembrane segments are predominantly hydrophobic, with a typical length around 20 amino acids. One can carry this description to the similarly 20 amino acid thick layers adjacent to cell membranes where proteins can interact most effectively, as they are temporarily confined to a narrower space. In this frontier surface space proteins can both interact on a 20 amino acid length scale, or evolve via exchange of modular elements of $\sim$20 amino acids. This is a post facto explanation for the remarkable discovery made by Ref. [14].

Instead of restricting the calculation of solvent-accessible amino acid areas to entire folded proteins from the PDB, Ref. [14] studied those areas associated with protein fragments centered on specific amino acids, of length L, $9 \leq L \leq 35$, an interval itself centered on the typical membrane thickness 21. (These fragments are sometimes called modules in models of protein evolution [16].) These solvent-accessible areas were then found to be linear in a log–log plot against L. This linearity is indicative of special properties of self-organized (nearly perfect) protein networks called self-similarity, and has been conjectured to be the key to a wide range of other nearly perfect networks. The amino-acid specific parameters of the model are universal fractals [17,18]. The membrane can be regarded as a catalytic substrate supporting protein–protein interactions in the frontier space.

This slightly long summary is incomplete, and several more steps are needed to exemplify bioinformatic scaling. Broadly speaking, the underlying idea is that a theory can be fruitful even if it is not exhaustive, providing it is over-determined by the fitted data. Bioinformatic scaling models contain usually only one adjustable chain segment length parameter W, and that length parameter is fixed by sequential features conserved over many species or strains. In the present case, because there are many cancers and many more cancer autoantibodies, we do not attempt to cover all possibilities exhaustively. Instead, we focus specifically on the landmark data of Ref. [8], and ask what bioinformatic scaling can tell us about the sensitive p53 epitopes they discovered.

The two bioinformatic scales used here are the modern MZ hydropathicity scale, based on protein conformational (thermodynamically second order) self-organized criticality of 5000 protein segments [14,19], and the β strand exposed (outside) residue amino acid propensity scale, based on a survey of nearly 2000 β strand structures by FTI [20]. Parallel calculations carried out with the thermodynamically first order KD standard hydropathicity scale, based on water–air protein unfolding [21], and the FTI β strand buried residue amino acid propensity scale [20], gave weaker results and are not reported, except for one example. Note that there is a fundamental topological similarity between the MZ and FTI scales, as both utilize a topological inside/outside dichotomy. The astonishing dimensional compressive power of such universal scales is displayed in the following multiple examples.

2. Materials and methods

How does p53 function as a universal tumor suppressor, the subject of nearly 10^5 papers? It acts as transcriptional activator, controlling the expression of a variety of genes important in cell cycle regulation and apoptosis [22]. Normally conformational changes in globular proteins, either evolutionary or mutational in origin, are best described with hydropathic scales Ψ, with β strand propensity a secondary physico-chemical factor [19,23]. Because we are interested in epitopes with sizes between 7- and 15-mers, we display in Fig. 1 $\Psi(n, 9)$ and $\Psi\ (n, 13)$, where $\Psi\ (n)$ is the MZ hydropathicity [14] of the nth amino acid in human p53, and superscale $\Psi(n, W)$ is $\Psi(n)$ smoothed by averaging over a sliding window of length W, appropriate to epitopes.