Insights in Biotechnology and Bioinformatics

Research Article

Folding and Binding Characteristics of Cancer - Associated Proteins

*Shi-Di Xiao and Bin-Guang Ma**

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China

Abstract

Cancer is an important disease that threatens human health. Protein mutations disrupt the normal proliferation regulation of cells, which leads to the formation of tumor. In this paper, we studied the characteristics of connectivity degree, folding rate, aggregation propensity of cancer-associated proteins and their relationships. We explored the aggregation propensity of neoantigen proteins in cancer immunotherapy and examined the overlap between neoantigen and cancer-associated proteins. The results showed that cancer-associated proteins have higher connectivity degree, larger folding rate, lower aggregation propensity and that there is a significant negative correlation between the folding rate and aggregation propensity of cancer-associated proteins. The aggregation propensity of neoantigen proteins are increased compared with the corresponding pre- mutated sequences. Our study helps to understand the protein folding and binding mechanisms and proffers a new indicator for the selection of neoantigens that are useful in designing immunotherapy for cancer treatment.

Keywords: Cancer; Protein interaction; Protein aggregation; Protein folding; Neoantigen

Introduction

More than a quarter of the people in the world end up with cancer [1], and many kinds of cancer cannot be cured effectively by ordinary surgery. A new strategy for tumor therapy has been proposed that is targeted immunotherapy based on neoantigen. Neoantigen is a protein that is mutated in tumor cells and causes an immune response. It is possible to effectively kill tumor cells by using T cells that can specifically recognize neoantigen [2]. There are also neoantigen-based personalized tumor vaccines that can effectively cure tumors [3,4]. The detection of correct and effective neoantigen is the key to this immunotherapy which is currently the biggest challenge, so exploring the characteristics of neoantigen to assist the detection of neoantigen has far-reaching significance.

Protein interactions are key factors to maintaining cell cycle, cell differentiation, and normal cell senescence, and disordered interaction networks can cause related diseases [5]. Cancer protein interaction networks and metabolic networks vary significantly from normal cells [6], and the mutations of cancer-associated proteins may be responsible for the disorder of their protein interaction networks. Folding rate is an important feature of protein that is associated with protein folding and complex formation processes. The characteristics of the folding rates of cancer-associated proteins remain largely unknown. Previous literature reports that hub proteins in the Protein-Protein Interaction (PPI) network of Escherichia coli K12 fold faster [7]. Although eukaryotes have larger complexity in alternative splicing and protein translation processes, human proteins in PPI networks may follow similar rules as well.

Protein aggregation is the process by which proteins are misfolded to form inactive or insoluble amyloid fibers [8]. A common form of protein aggregation is amyloid and amorphous, and the beta-sheet structure formed during protein folding can lead to insoluble amyloid structures [9]. Protein aggregation and precipitation can cause a variety of diseases, such as Alzheimer's, Parkinson's, spongiform encephalopathy, and type 2 diabetes [10]. Cancer is also a disease highly associated with abnormal aggregation of proteins [11], and studies have found protein aggregation caused by the misfolding of P53 tumor suppressor protein in cancer cells [12]. The degree of protein aggregation may affect the ability of the immune system to recognize tumor cells.

In this paper, we explored the characteristics of folding rate, aggregation propensity and connectivity degree in the human PPI network of cancer-associated proteins as well as their interrelationships. It was found that protein folding process has effect on protein aggregation. Our results showed that cancer-associated proteins have greater connectivity in the PPI network and the folding rates of hub proteins in the network are also larger. In addition, the characteristics of aggregation propensity of neoantigens in tumor cells were analyzed, which can serve as a feature to identify neoantigen.

Materials and Methods

Data source of cancer-associated protein sequences and interactions

The human cancer gene data were downloaded from NCI Genomic Data Commons [13] and 574 cancer genes were identified. To ensure reliability, only 464 cancer genes with SSM affected cases in cohort ≥ 1% were used. The corresponding cancer protein sequences were obtained from the Uniprot database [14]. The human protein interaction data came from the HIPPIE database [15] and a total

Citation: Xiao SD, Ma BG. Folding and Binding Characteristics of Cancer - Associated Proteins. Insights Biotechnol Bioinforma. 2019; 1(1): 1001.

Publisher Name: Medtext Publications LLC

Manuscript compiled: May 16th, 2019

***Corresponding author:** Bin-Guang Ma, Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China, Tel: 862787280877; E-mail: mbg@mail.hzau.edu.cn

of 16,688 protein sequences and the corresponding interaction information were obtained. The connectivity degree of a protein in the human PPI network is defined as the number of its interacting proteins. The proteins with the top x% (x=10, 15, 20, 25) highest connectivity degrees are defined as hub proteins in the PPI network.

Connectivity degree and folding rate comparisons

In the HIPPLE database, only 461 of the 464 cancer-associated proteins have interaction information. The connectivity degrees of the 461 cancer-associated proteins and 16,920 non-cancer proteins were compared using two-sample t-tests. In the HIPPLE database, only 16,688 proteins have their amino acid sequence information. Among the 16,688 proteins from the HIPPLE database, the proteins with the top x% (x=10, 15, 20, 25) highest connectivity degrees are assigned as hub proteins in the human PPI network and the others non-hub proteins. The folding rates for these proteins were calculated from the amino acid sequences by using our CI indicator [16] and comparison between hub and non-hub proteins was performed using a two-sample t-test.

Comparisons of folding rate and aggregation propensity

The folding rates and aggregation propensities of proteins were calculated from the amino acid sequence by using our CI [16] and AP [17] indicators. To exclude the effects of amino acid chain length, 464 out of 16,224 non-cancer proteins with amino acid sequence were randomly sampled with an average length the same as that of the cancer- associated proteins. The folding rates and aggregation propensity of 464 cancer- associated proteins were calculated and compared with those of non-cancer proteins using a two-sample t-test. The correlation between folding rate and aggregation propensity was assessed using the Pearson correlation coefficient.

Features of neoantigen

The detailed information of neoantigen of 16 tumor tissues was downloaded from TSNAdb [18]. Binding level is measured by the binding strength between the neoantigen peptide and the HLA (Human Leukocyte Antigen) allele. Neoantigens with a binding level of "strong binding" after mutation were selected as the trusted neoantigens, and the neoantigen proteins are sorted by the occurrence frequency. The aggregation propensities of neoantigen proteins before and after mutation were calculated and compared in tumor tissues. In addition, the overlap between the cancer-associated proteins and the neoantigen proteins was also analyzed.

Results

Cancer-associated proteins have higher connectivity degrees in the human PPI network

The performance of many biological functions relies on complex PPI networks whose malfunction can cause various diseases [19]. In the human PPI network, the protein nodes with high connectivity tend to have more important cellular functions [20]. We found that the degree distribution of cancer-associated proteins in the human PPI network is significantly different from that of non-cancer proteins (Figure 1A). The average connectivity degree of cancer-associated proteins is significantly (t-test: p<2.2E-16) higher than that of non-cancer proteins (Table 1). It was found that cancer-associated proteins widely regulate cell growth and development processes, ensuring normal proliferation and apoptosis of cells [21]. Once the corresponding genes are mutated, the proteins cannot function normally, and the cell proliferation is out of control, resulting in tumors.

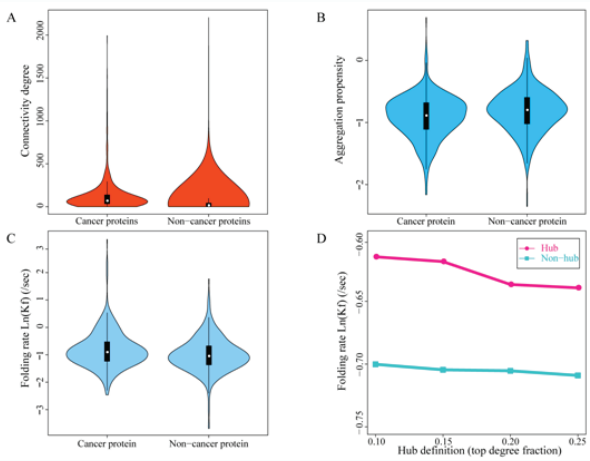

Figure 1: (A) Connectivity degree comparison between cancer-associated and non-cancer proteins in human. Cancer-associated proteins have a higher degree of connectivity than non-cancer proteins. (B) Aggregation propensity comparison between cancer-associated and non-cancer proteins. Cancer-associated proteins have a lower aggregation propensity than non-cancer proteins. (C) Folding rate comparison between cancer-associated and non-cancer proteins. Cancer-associated proteins are folding faster than non-cancer proteins. (D) Folding rate comparison between hub and non-hub proteins in the human PPI network. Hub proteins are folding faster than non-hub proteins with different definition criteria of hub proteins.

Table 1: t-test information for the comparisons of connectivity degree, folding rate and aggregation propensity between cancer-associated and non-cancer proteins.

Mean (Number)	Cancer protein	Non-cancer protein	t-statistic/p-value
Connectivity degree	126.12 (461)	36.58 (16920)	9.86/<2.2E-16
Aggregation propensity	-0.91 (464)	-0.81 (464)	-4.22/<2.67E-05
Folding rate	-0.79 (464)	-0.96 (464)	4.17/3.35E-05

Cancer-associated proteins have lower aggregation propensity

In addition, we calculated the aggregation propensities of cancer-associated proteins using our AP indicator [17] and compared with those of the non-cancer proteins. The results of comparison showed that the aggregation propensities of cancer-associated proteins were lower than those of non-cancer proteins (Figure 1B), and the difference between the average values is significant (Table 1). This result indicates that cancer- associated proteins are not easy to form precipitated polymers, and can interact with other proteins more broadly and efficiently, thereby ensuring normal cell proliferation.

Cancer-associated proteins and hub proteins are folding faster

The folding rates of cancer-associated and non-cancer proteins were predicted by using FDserver [22] and shown in violin plot (Figure 1C). Compared by two-sample t-test, it was shown that cancer-associated proteins have significantly (p-value=3.35E-05) larger folding rates than non-cancer proteins averagely (Table 1). As shown above, cancer- associated proteins have larger connectivity degree and thus they are of higher probability of being hub proteins. It was reported that the hub proteins in the PPI network of Escherichia coli K12 have relatively larger folding rates [7]. We explored whether

this phenomenon is also present in human PPI network. The hub proteins here are defined as the proteins with the top x% (x=10, 15, 20, 25, respectively) highest connectivity degree. FDserver [22] was used to predict the folding rate of hub and non-hub proteins and the average folding rates of hub and non-hub proteins were plotted against the hub definition criterion (top degree faction). As shown in Figure 1D, the average folding rates of hub proteins are always larger than non-hub proteins on the whole definition range. Two-sample t-tests indicate that the differences are significant (Table 2). These results are consistent with the case in the bacteria Escherichia coli K12 [7]. The results indicate that hub proteins interact with more protein partners and may require higher folding rates to meet their functional demands.

Protein folding rate and aggregation propensity are negatively correlated

Protein folding and aggregation are processes in which proteins form a tertiary structure from a primary structure (amino acid sequence) and the relationship between these two processes are still unclear. We calculated the protein folding rates and aggregation propensities of 464 cancer-associated based on their amino acid sequences and found a significant negative correlation: Pearson Correlation Coefficient (PCC)=-0.61, p-value<2.2E-16 (Figure 2), which indicates that faster folders have smaller aggregation propensity. It is understandable in terms of the β-sheet conformation, because β-sheets are of an important influential factor for both protein folding and aggregation. It has been reported that the content of β-sheet is negatively correlated with folding rates [23], while positively correlated with aggregation propensities of proteins [17,24]. The negative correlation between the protein folding rate and aggregation propensity indicates that slower folding proteins may expose the β-sheets for a longer time and therefore promote protein aggregation.

Features of neoantigen

Neoantigen is now a new strategy for cancer immunotherapy and how to obtain reliable neoantigen is a key challenge to achieving good effects [25]. The specific neoantigen will stimulate the T cell immune response which can specifically and effectively kill cancer cells by enriching the corresponding T cells. We used the protein sequence mutation information of neoantigen to detect the change in the aggregation propensity of the neoantigens in each tumor tissues. The results indicate that the aggregation propensity of neoantigen is generally increased in tumor tissues (Figure 3). Neoantigens have higher aggregation propensity than pre-mutated protein sequences. In addition, the numbers of neoantigens overlapping with cancer-associated proteins in each cancer tissue is more than random, and the distribution of cancer-associated proteins in the top 100 neoantigens is significantly higher than the random distribution by Fisher tests (Table 3), showing the association between neoantigen and cancer-associated proteins. Although the aggregation propensities of some neoantigen proteins are lower in the tumor specific mutations, most of them are higher and thus the overall trend is increasing (see Supplementary Data for details). The results indicate that the higher aggregation propensity of the mutated proteins in tumors cells makes easier protein precipitation which brings greater chance to be recognized by T cells to cause an immune reaction, so aggregation propensity may serve as an indicator for the probability of protein mutations being neoantigen which can be used for cancer immunotherapy.

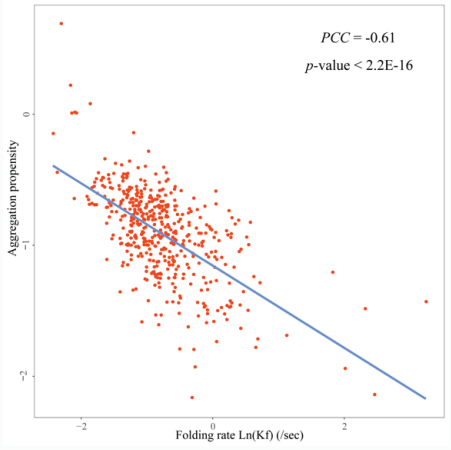

Figure 2: Correlation between folding rates and aggregation propensities in human cancer-associated proteins. A significant negative correlation (PCC=-0.61, p-value <2.2E-16) is observed.

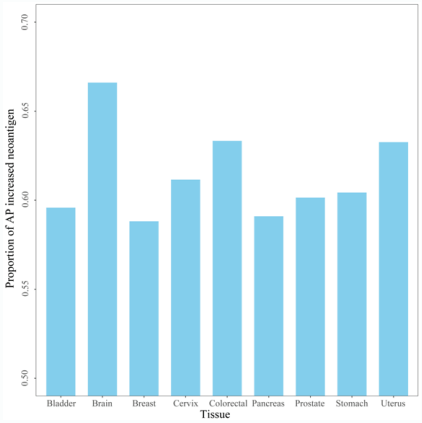

Figure 3: Proportion of aggregation-propensity-increased neoantigens in all neoantigens for each cancer tissues. More than half of the neoantigens' aggregation propensities increased compared with pre-mutated sequences.

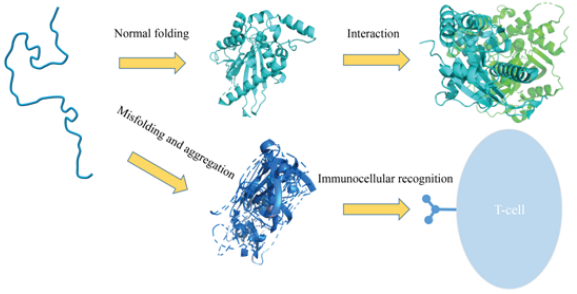

Figure 4: Consequences of normal and abnormal protein folding.

Table 2: t-test information for the comparisons of folding rates between hub and non-hub proteins.

Top degree fraction	Hub Number	Non-hub number	Hub average folding rate	Non-hub average folding rate	t-statistic/p-value
10%	1669	15019	-0.61	-0.70	-4.16/3.23E-5
15%	2503	14185	-0.62	-0.70	-4.87/1.18E-6
20%	3338	13350	-0.64	-0.71	-4.39/1.13E-5
25%	4172	12516	-0.64	-0.71	-4.82/1.43E-6

Table 3: Overlap between neoantigen and cancer-associated proteins.

	Bladder	Kidney	Colorectal	Uterus	Stomach	Breast	Brain	Cervix
Neoantigen proteins	10985	6055	14189	17216	12073	10188	4314	8704
Neoantigen & Cancer proteins	360	234	406	453	381	348	172	306
Top100 p-value[a]	5.44E-08	9.42E-06	8.05E-05	0.002	0.002	0.004	0.004	0.005

[a]Fisher test p-values for the expected and actual numbers of cancer-associated proteins in the top100 neoantigen proteins.

Discussion

In this paper, analysis of the human PPI network indicated that cancer-associated proteins interact with significantly more proteins than non-cancer proteins, indicating that cancer- associated proteins are widely involved in protein regulation, similar to previously reported results [26]. At the same time, we found that the folding rates and aggregation propensities of cancer-associated proteins and non-cancer-associated proteins are significantly different. For cancer-associated proteins, the folding rates were significantly larger while the aggregation propensity was significantly smaller.

Protein folding and binding have been a hotspot for studying protein functions [27,28]. The link between protein folding rates and connectivity degrees in the PPI network is important for a deeper understanding of protein folding and binding mechanisms. The results in this paper showed that the folding rates of hub proteins in the human PPI network is larger, consistent with the case in prokaryotes [7], which may be related to the corresponding protein functions [29]. Studies have also shown that evolutionary processes prefer to select proteins with fast folding speed [30]. The connection between protein folding rates and connectivity degrees in the human PPI network may also reflect some selection advantages in evolution that is ubiquitous in both eukaryotes and prokaryotes.

We also found that there is a significant negative correlation between the folding rate and aggregation propensity for cancer-associated proteins, namely, the slower the folding, the stronger the aggregation. The aggregation of P53 protein in cancer cells [31], may be due to a decrease in the folding rate of p53 protein after mutation, resulting in increased exposure time of the β-sheets structure [32] which increases the likelihood of protein aggregation. There may be two cases in the process of protein folding: a protein folds normally and fulfills its function, or a protein fold abnormally forming aggregates. In the latter case, the protein aggregates may be recognized and cleared by immune cells. However, when the immune system does not respond in time, it will have serious consequences (Figure 4). Our finding may bring about a new strategy for the treatment of cancer, i.e., designing new molecular chaperones to assist the cancer-associated proteins to complete normal folding.

Another important finding is that the aggregation propensities of neoantigens are significantly increased compared with pre-mutated sequences. It can be considered to use the protein aggregation propensity as a feature for the detection of neoantigen, saving the cost of experiments and assisting the screening of neoantigen [33]. There is a model for assessing the fitness of neoantigens [34], which can predict the therapeutic effects of neoantigens. Adding aggregation propensity as a feature to the model may improve the predictive power and promote the development of personalized cancer treatment. In addition, by examining the overlap between neoantigen and cancer-associated proteins, we found that neoantigens are significantly enriched with cancer-associated proteins, showing their connection. The mutations of cancer-associated proteins affect the normal life activities of cells, and it is these cancer-associated proteins that have a high probability of stimulating an immune response and have the potential to become neoantigen.

In this paper, the correlations between the folding rates, aggregation propensity and connectivity degree of cancer-associated proteins were analyzed, and the aggregation propensity characteristics of neoantigens were investigated. Our study will be helpful to the understanding of protein folding and binding mechanisms and to the design of cancer immunotherapy based on neoantigens.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant 31570844), Project 2662016PY094 supported by the Fundamental Research Funds for the Central Universities. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015;65(1):5-29.
2. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science. 2014;344(6184):641-5.
3. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature. 2017;547(7662):217-21.
4. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017;547(7662):222-6.
5. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I. Fundamentals of protein interaction network mapping. Mol Syst Biol. 2015;11(12):848.
6. Dong W, Keibler MA, Stephanopoulos G. Review of metabolic pathways activated in cancer cells as determined through isotopic labeling and network analysis. Metab Eng. 2017;43(Pt B):113-24.
7. Xu HR, Cheng JF, Hu XP, Chu YY, Ma BG. Are protein hubs faster folders? Exploration based on Escherichia coli proteome. Amino Acids. 2016;48(12):2747-53.
8. Hipp MS, Park SH, Hartl FU. Proteostasis impairment in protein- misfolding and-aggregation diseases. Trends Cell Biol. 2014;24:506-14.
9. Fandrich M. On the structural definition of amyloid fibrils and other polypeptide aggregates. Cell Mol Life Sci. 2007;64(16):2066-78.
10. Dobson CM. The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci. 2001;356(1406):133-45.
11. Costa DC, de Oliveira GA, Cino EA, Soares IN, Rangel LP, Silva JL. Aggregation and prion-like properties of misfolded tumor suppressors: is cancer a prion disease?. Cold Spring Harb Perspect Biol. 2016;8(10).
12. Wilcken R, Wang G, Boeckler FM, Fersht AR. Kinetic mechanism of p53 oncogenic mutant aggregation and its inhibition. Proc Natl Acad Sci U S A. 2012;109(34):13584-9.
13. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109-12.
14. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46(5):2699.
15. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Res. 2017;45:D408-14.

16. Ma BG, Guo JX, Zhang HY. Direct correlation between proteins' folding rates and their amino acid compositions: an ab initio folding rate prediction. Proteins. 2006;65(2):362-72.

17. Ma BG, Goncearenco A, Berezovsky IN. Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. Structure. 2010;18(7):819-28.

18. Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNAdb: A database for tumor-specific neoantigens from immunogenomics data analysis. Genomics Proteomics Bioinformatics. 2018;16(4):276-82.

19. Liu W, Sun Z, Xie H. The analyses of human inherited disease and tissue- specific proteins in the interaction network. J Biomed Inform. 2016;61:10-8.

20. Zhang M, Zhu C, Jacomy A, Lu LJ, Jegga AG. The orphan disease networks. Am J Hum Genet. 2011;88(6):755-66.

21. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011;11(2):85-95.

22. Ma BG, Chen LL, Zhang HY. FDserver: A web service for protein folding research. Nature Precedings. 2008.

23. Ji-Tao H, Jin-Pei C, Hui C. Secondary structure length as a determinant of folding rate of proteins with two- and three-state kinetics. Proteins Structure Function and Bioinformatics. 2010;67(1):12-7.

24. Ferrolino MC, Zhuravleva A, Budyak IL, Krishnan B, Gierasch LM. Delicate balance between functionally required flexibility and aggregation risk in a beta- rich protein. Biochemistry. 2013;52(49),8843-54.

25. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. 2016;8:11.

26. Cheng F, Jia P, Wang Q, Lin CC, Li WH, Zhao Z. Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. Mol Biol Evol. 2014;31(8):2156-69.

27. Gianni S, Dogan J, Jemth P. Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics?. Curr Opin Struct Biol. 2016;36:18-24.

28. Muñoz V, Campos LA, Sadqi M. Limited cooperativity in protein folding. Curr Opin Struct Biol. 2016;36:58-66.

29. Giri Rao VV, Gosavi S. Using the folding landscapes of proteins to understand protein function. Curr Opin Struct Biol. 2016;36:67-74.

30. Debès C, Wang M, Caetano-Anollés G, Gräter F. Evolutionary optimization of protein folding. PLoS Comput Biol. 2013;9(1):e1002861.

31. Kim S, An SS. Role of p53 isoforms and aggregations in cancer. Medicine (Baltimore). 2016;95(26):e3993.

32. Francioso A, Punzi P, Boffi A, Lori C, Martire S, Giordano C, et al. ß-sheet interfering molecules acting against beta-amyloid aggregation and fibrillogenesis. Bioorg Med Chem. 2015; 23(8):1671-83.

33. Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long- term survivors of pancreatic cancer. Nature. 2017;551(7681):512-6.

34. Luksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature. 2017;551(7681):517-520.