FEBS Letters 587 (2013) 136–141

journal homepage: www.FEBSLetters.org

Molecular dynamics simulations and statistical coupling analysis reveal functional coevolution network of oncogenic mutations in the CDKN2A–CDK6 complex

Jingwen Wang [b,1], Yuqi Zhao [a,1], Yanjie Wang [c,1], Jingfei Huang [a,d,*]

[a] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Eastern Jiaochang Road, Kunming, Yunnan 650223, China
[b] Yantai Yuhuangding Hospital, Yantai, Shandong Province 264000, China
[c] Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[d] Kunming Institute of Zoology, Chinese University of Hongkong, Joint Research Center for Bio-resources and Human Disease Mechanisms, Kunming 650223, China

ARTICLE INFO

Article history:
Received 5 September 2012
Revised 19 October 2012
Accepted 6 November 2012
Available online 21 November 2012

Edited by Takashi Gojobori

Keywords:
Inter-protein coevolution
Oncogenic mutations
Melanoma
Molecular dynamics simulation
CDK6–CDKN2A complex

ABSTRACT

Coevolution between proteins is crucial for understanding protein–protein interaction. Simultaneous changes allow a protein complex to maintain its overall structural–functional integrity. In this study, we combined statistical coupling analysis (SCA) and molecular dynamics simulations on the CDK6–CDKN2A protein complex to evaluate coevolution between proteins. We reconstructed an inter-protein residue coevolution network, consisting of 37 residues and 37 interactions. It shows that most of the coevolved residue pairs are spatially proximal. When the mutations happened, the stable local structures were broken up and thus the protein interaction was decreased or inhibited, with a following increased risk of melanoma. The identification of inter-protein coevolved residues in the CDK6–CDKN2A complex can be helpful for designing protein engineering experiments.

1. Introduction

Cyclin-dependent kinase inhibitor 2A (CDKN2A) acts as a negative regulator of the proliferation of normal cells by interacting strongly with Cyclin-dependent kinases, such as CDK6, inhibiting their ability to interact with cyclins D and to phosphorylate the retinoblastoma protein [1]. Candidate-gene studies and genome-wide association studies (GWAS') for melanoma and the melanoma-associated phenotypes have identified many variants in CDKN2A associated with melanoma risk in the general population [2–4]. According to reports to date, mutations in CDKN2A account for approximately 40% of familial melanoma cases [5]. High cancer susceptibility determines that amino acid substitutions in CDKN2A may be restricted. However, surprisingly, the protein sequence of CDKN2A is not evolutionarily conserved. For example, the sequence identity of human CDKN2A protein with its mouse ortholog is 69.23%.

Previous studies reported that protein–protein interactions effectively mediate molecular function and they are the result of specific interactions between protein interfaces [6,7]. In general, functional constraints restrict amino acid substitutions in protein sequences. The results can be detected in multiple sequence alignments as evolutionary fingerprints. For example, correlated mutations (or coevolution in another word) of the residues at two distinct alignment positions, is a result of functional constraints that force compensating mutations for specific residue changes [8]. With the increasing number of genomes and meta-genomes sequenced, the correlated mutation analyses on intra-protein amino acids have been extensively performed [8–10]. However, studies on inter-protein coevolution are lacking. Referring to CDKN2A, communications between inter-protein amino acids at the binding interface of CDK6–CDKN2A complex may define its biological role, such as inhibitor specificity and allosteric regulation. Besides, CDKN2A and CDK6 are particularly suitable for correlated mutation analysis because the large sets of sequences are available and the crystal structure of the protein complex has been resolved [11].

In this study, we perform an accurate detection of functionally coevolved residue network of CDK6–CDKN2A complex using phylogenetic information and molecular dynamic simulations. First,

* Corresponding author at: State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Eastern Jiaochang Road, Kunming, Yunnan 650223, China. Fax: +86 0871 5199200.
E-mail address: huangjf@mail.kiz.ac.cn (J. Huang).
[1] These authors contributed equally to this work.

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.11.001

we collected homologue sequences of human CDK6 and CDKN2A proteins and then concatenated them according to species. Then, the concatenated sequences were aligned and submitted to statistical coupling analysis (SCA) [12] for detecting the inter-protein coevolved residues. At last, we estimated the effects of mutations on CDKN2A using molecular dynamics simulations. We also discussed the applications of our results.

2. Materials and methods

2.1. Data source and multiple sequence alignment

The sequences of both CDK6 and CDKN2A were collected from the non-redundant database of protein sequences by NCBI BLASTP 2.2.25 (with default settings except *E*-value = 0.01). The collected sequences were aligned using the multiple sequence alignment program, MUSCLE [13], which processed all sequences at the same time to prevent inconsistencies.

2.2. Correlated mutation analysis

The static energy for each site and statistical coupling energy between any two sites were calculated by using the statistical coupling analysis (SCA) method [12].

To perform SCA, we manually adjusted the aligned sequences described above to improve alignment in less conserved positions. The conservation criterion ΔG^{stat} for the multiple sequence alignment (MSA) is defined as follows:

$$\Delta G_i^{\text{stat}} = \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \tag{1}$$

where P_i^x is the binomial probability of finding a given residue x in the i position in the MSA and P_{MSA}^x is the binomial probability of finding the residue x in the MSA. Following the improvements of the method [14], we computed the frequencies of finding each residue in the MSA directly from the alignment of CDK6 and CDKN2A. A minimal size in the data set for perturbation experiments was selected to guarantee the statistical equilibrium. For this purpose, we computed averages of ΔG^{stat} values for the ten less conserved positions and stepwise reduced the working data set by randomly excluding the sequences. Analysis of the average of ΔG^{stat} for the least conserved positions versus data set size defined a minimal size for the subset for $\Delta\Delta G^{\text{stat}}$ calculations to be about half of the total alignment size (80 sequences). Perturbations were performed in every sequence position that fitted the latter criterion for the subset definition. The $\Delta\Delta G^{\text{stat}}$ was computed as follows:

$$\Delta\Delta G_{ij}^{\text{stat}} = \sqrt{\sum_x \left(\ln \frac{P_{i|\delta j}^x}{P_{\text{MSA}|\delta j}^x} - \ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \tag{2}$$

where $P_{i|\delta j}^x$ is the binomial probability of finding residue x in the i position in the subset of the alignment chosen by the perturbation in the j position. The final matrix containing all the performed perturbations was submitted to iterative cycles of cluster analysis in MATLAB. After each cycle, positions with weak signals were discarded. Finally, the alignment of 159 sequences including 427 sites was available for analysis. The SCA correlation matrix between amino acids was turned into *Z*-scores (also called Standard scores). If a *Z*-score was above a fixed threshold (cutoff = 4), two corresponding sites were linked by an edge, and each site was represented as a node.

2.3. Molecular dynamic simulations

The crystal structure of CDK6–CDKN2A complex was retrieved from Protein Data Bank (http://www.rcsb.org/pdb/) with the accession (PDB ID: 1BI7) [11]. The structure was then solvated in a TIP3P water box [15] and ionized by NaCl (0.152 M) to mimic physiological conditions. Mutant models of CDK6–CDKN2A complex were made using MODELER program [16].

All simulations were performed using NAMD 2.8 [17] and the CHARMM31 force field with CMAP correction [18,19]. The ionized systems were minimized for 50,000 integration steps and constrained equilibrated for 10 ns with 2 fs time stepping and frames stored each picosecond. Constant temperature (T = 310 K) was enforced using Langevin dynamics with a damping time constant of 5 per picosecond. Constant pressure (p = 1 atm) was enforced through the Nosé-Hoover Langevin piston method with a decay period of 100 fs and a damping time constant of 50 fs. Van der Waals interaction cutoff distances were set at 12 Å (smooth

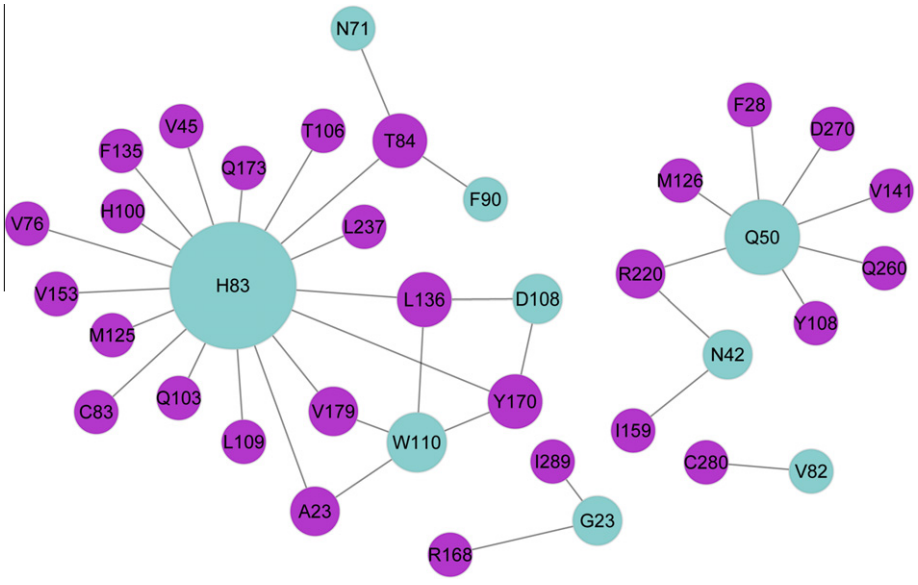

Fig. 1. Reconstruction of coevolution network of inter-protein residues of CDKN2A–CDK6 complex. The nodes represent residues in proteins (CDKN2A in green and CDK6 in red) while the links between them are the coevolution relationships.