

# Lead discovery and in silico 3D structure modeling of tumorigenic FAM72A (p17)

Subrata Pramanik · Arne Kutzner · Klaus Heese

Received: 3 August 2014 / Accepted: 9 September 2014 / Published online: 20 September 2014  
© International Society of Oncology and BioMarkers (ISOBM) 2014

**Abstract** FAM72A (p17) is a novel neuronal protein that has been linked to tumorigenic effects in non-neuronal tissue. Using state of the art in silico physicochemical analyses (e.g., I-TASSER, RaptorX, and Modeller), we determined the three-dimensional (3D) protein structure of FAM72A and further identified potential ligand-protein interactions. Our data indicate a Zn<sup>2+</sup>/Fe<sup>3+</sup>-containing 3D protein structure, based on a 3GA3\_A model template, which potentially interacts with the organic molecule RSM ((2s)-2-(acetylamino)-N-methyl-4-[(R)-methylsulfinyl] butanamide). The discovery of RSM may serve as potential lead for further anti-FAM72A drug screening tests in the pharmaceutical industry because interference with FAM72A's activities via RSM-related molecules might be a novel option to influence the tumor suppressor protein p53 signaling pathways for the treatment of various types of cancers.

**Keywords** Ugene · FAM72A · In silico · 3D structure · Cancer · RSM

**Electronic supplementary material** The online version of this article (doi:10.1007/s13277-014-2620-7) contains supplementary material, which is available to authorized users.

S. Pramanik · K. Heese (✉)  
Graduate School of Biomedical Science and Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Republic of Korea  
e-mail: klaus.heese@rub.de

K. Heese  
e-mail: klaus@hanyang.ac.kr

A. Kutzner  
Department of Information Systems, College of Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Republic of Korea

## Introduction

Cancer is still a major health threat to people all over the world, including to the Korean population. In total, about 200,000 cancer cases and 70,000 cancer deaths occur annually in Korea, while more than 1,000,000 prevalent cancer cases are identified in Korea's population today. Notably, the overall cancer incidence in Korea increases rapidly with 70,000 deaths occurring annually while the incidence of all cancers combined shows an approximate annual increase of 3.5 % [1].

The experimental molecular, cellular, and bioinformatics approach employed in this study has been used previously to identify novel oncogenes with therapeutic potential. Using these methods, we discovered the new membrane-bound FAM72A protein (also known as p17, Ugene, or LMPIP) with demonstrated highly promising clinical relevance to survival/death outcomes in patients with various kinds of cancers, as it can be linked to tumorigenic effects in non-neuronal tissues [2–4]. Our group also unraveled FAM72A's mode of action, and our data demonstrate that FAM72A interacts with various tumor suppressor proteins that are (epi-) genetically modified in cancer [3] and thus, interference with FAM72A's activities might be a novel option to influence the tumor suppressor protein p53 signaling pathways for the treatment of tumors.

Upon identification of a potential therapeutic target, the pivotal challenge of current cancer research is the resolution of the target's three-dimensional (3D) protein structure for the application of high-throughput drug-screening tests [5]. Knowledge of a protein's structure is important for the general understanding of a protein's function, and it is particularly essential for the development of target-specific drugs [6]. Unfortunately, laboratory-based drug development takes much time, consumes huge amounts of money, and frequently involves ethical issues regarding animal experiments [7, 8]. Thus, in silico studies are now becoming more important than ever to push forward cancer research for the development of

novel drugs [9–11]. In the present study, we used a state-of-the-art approach to resolve FAM72A's 3D protein structure by multiple (non-)template modeling approach analyses [12–16] and identified a novel potential chemical molecule that may serve as lead for screening tests in search of novel drugs for the treatment of FAM72A-based cancers.

## Methods

### FAM72A reference sequence

The sequence of the human FAM72A protein (gene ID, 729533, NP\_001116640.1, FAM72A) consists of 149-amino acids (AAs) and was retrieved from the National Center for Biotechnology Information (NCBI) at <http://www.ncbi.nlm.nih.gov/protein> (Supplementary Fig. S1).

### Primary template and structure prediction using NCBI's protein data bank search

NCBI's protein data bank (PDB) database was extensively screened for appropriate template selection using BLAST, PSI-BLAST, and DELTA-BLAST to find the most suitable homologous 3D structure for FAM72A. The NCBI-PDB search was done according to the general strategy as briefly outlined at: [http://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb\\_how\\_to\\_search\\_by\\_gene.html](http://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb_how_to_search_by_gene.html). The DELTA-BLAST tool detects distant homologs in a protein database search and provides better quality alignments than do BLAST and PSI-BLAST. DELTA-BLAST searches a database of pre-constructed position-specific scoring matrices (PSSMs) before searching a protein-sequence database, to achieve better homology detection [17]. Details of homology sequence alignments are shown in the Supplementary data material (Supplementary Figs. S2–S4).

### Primary template and protein structure prediction using the online server-based bioinformatics tools

In addition to using NCBI's PDB for searching for 3D structures of homologous protein, we applied various online-based structure prediction tools such as (i) Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>), (ii) 3DJIGSAW v.3.0 (<http://bmm.cancerresearchuk.org/~3djigsaw/>), and (iii) Swiss Model (<http://swissmodel.expasy.org/>) to develop a 3D structure and find a suitable template for 3D FAM72A protein structure modeling (Supplementary Figs. S5–S20).

### Primary template and structure prediction using RaptorX

RaptorX is a protein structure prediction server for 3D protein structures of protein sequences without close homologs in the

PDB [14]. Given an input sequence, RaptorX predicts its secondary and tertiary structures as well as solvent accessibility and disordered regions. The basic concept is shown in the Supplementary data material (Supplementary Fig. S21).

### Primary template and structure prediction using multiple templates with I-TASSER

A standard rational protein threading method builds the 3D protein structure of a target protein sequence using a single template protein. Although many experimentally obtained protein structures are deposited in the NCBI-PDB, it is still far from being a comprehensive human 3D proteome structure database. Because of the limited number of proteins in the database, any given human target protein is likely to receive only a fairly poor match with any single-solved 3D protein structure in the PDB, which could be used as a potential starting template to predict the 3D protein structure. Consequently, since our FAM72A protein had only a relatively remote homology with a number of templates that were suggested by the NCBI-PDB database, Phyre2, 3DJIGSAW v.3.0, or Swiss Model, we needed to extend the classical protein threading method so that a target protein sequence could be threaded onto multiple templates simultaneously and thus, its 3D protein structure model be built from multiple template structures. In this context, we also implemented the multiple templates/threading protocol using the I-TASSER server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>). The I-TASSER server is an online platform for protein structure and function predictions [13, 18]. 3D models are built based on multiple-threading alignments by Local Meta-Threading Server (LOMETS) and iterative template fragment assembly simulations. LOMETS is an online Web service for protein structure prediction. It generates 3D models by collecting high-scoring target-to-template alignments from nine locally installed threading programs (FFAS, HHsearch, MUSTER, PPA, PRC, PROSPECT2, SAM-T02, SP3, and SPARKS). The reliability of the structure prediction results are reported as a template modeling (TM)-score that is a number between 0 and 1. A TM-score <0.17 indicates a random model, while a TM-score >0.5 corresponds to two structures of similar topology (Supplementary Fig. S22).

### Protein structure prediction with Modeller v.9.13

Our preanalyzed suggested templates from the NCBI-PDB database, Phyre2, 3DJIGSAW v.3.0, and Swiss Model were subjected to comparative modeling using the Modeller 9.13 software as described [19, 20] and applied previously [21–33]. Additionally, the best suggested templates from RaptorX were also applied to the comparative modeling analysis using Modeller 9.13 software.

### Protein structure evaluation

The stereochemical qualities of the obtained FAM72A models were ultimately evaluated with the program PROCHECK in order to prove the structural quality of the *in silico* 3D protein structures and for selection of the best model for further detection of potential ligand-binding sites [34].

### Protein structure comparison analysis by alignment

The superimposition or 3D alignment of 3D protein structures is an important method to evaluate the common 3D substructure of a set of molecules. Structure alignments have been done using TM-align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>), which is a highly optimized algorithm for protein structure comparison and alignment [35]. For two protein structures of unknown equivalence, TM-align first generates the residue-to-residue alignment based on structural similarity using dynamic programming iterations. An optimal superposition of the two structures, as well as the TM-score value, which scales the structural similarity, is then returned. The TM-score value lies between 0 and 1. In general, a TM-score <0.2 indicates that there is no similarity between two structures while a TM-score >0.5 means the structures share the same SCOP/CATH fold [36]. Additionally, root-mean-square deviation (RMSD) is also checked to determine the structural similarity of 3D structure alignments. The most commonly used metric is the RMSD, in which the root-mean-square distance between corresponding residues is calculated after an optimal rotation of one structure to another [37, 38]. Since the RMSD weights the distances between all residue pairs equally, a small number of local structural deviations could result in a high RMSD, even when the global topologies of the compared structures are similar. Furthermore, the average RMSD of randomly related proteins depends on the length of compared structures, which renders the absolute magnitude of RMSD meaningless [39]. In connection with the 3D structure alignment, additional sequence alignments with different matrices (PAM-30, PAM-70, PAM-250, BLOSUM-80, BLOSUM-62, BLOSUM-45, BLOSUM-50, and BLOSUM-90) were carried out for the templates suggested from the NCBI-PDB database, Phyre2, 3DJIGSAW, and Swiss Model based on the FAM72A sequence in order to confirm the overall sequence similarity.

### Lead discovery by protein-ligand-binding-site prediction

Since FAM72A plays a significant role in tumorigenesis [3, 4], the present study on the theoretical 3D protein structure modeling of FAM72A led us to search for a potential ligand-binding site of FAM72A as the identification of protein-ligand-binding sites is a decisive pathway for the elucidation of protein function and development of therapeutics. A meta-

server approach (e.g., via <http://zhanglab.ccmb.med.umich.edu/>) using COACH has been used to identify a predicted potential FAM72A protein-ligand-binding site [40–42]. Starting from a given 3D structure of target proteins, COACH generates complementary ligand-binding-site predictions using two comparative methods, TM-SITE and S-SITE, which recognize ligand-binding templates from the BioLiP protein function database (<http://zhanglab.ccmb.med.umich.edu/BioLiP/index.html>) by binding specific substructure and sequence profile comparisons. These predictions were combined with results from other methods (including COFACTOR, FINDSITE, and ConCavity) to generate final ligand-binding-site predictions. The reliability of the ligand-protein-binding-site prediction was checked by confidence score (C-score). The C-score is in the range of zero to one (0–1), where a C-score of higher value signifies a model with a high confidence and vice versa [18].

## Results

### Primary template and structure prediction using NCBI's PDB search

Applying NCBI's protein-BLAST homology search, we found three homologous protein 3D structure suggestions in the PDB database for FAM72A (149 AAs): (i) 1YQ3\_D (103 AAs), (ii) 4OGC\_A (1101 AAs), and (iii) 4OGE\_A (1101 AAs) (Supplementary Figs. S2–S4). The selected templates were then forwarded for the Modeller analysis.

### Primary template and structure prediction using Phyre2

Phyre2 suggested the 3GA3\_A (133 AAs) template as one of the best homologous templates for a possible 3D FAM72A protein structure (Supplementary Fig. S5). Phyre2 predicted the 3D FAM72A (149 AAs) protein structure (with 128 AAs) based on the 3GA3\_A template (133 AAs) (Supplementary Fig. S6). Data obtained were prechecked by various means such as Ramachandran plot and others (Supplementary Figs. S6–S7). The selected template 3GA3\_A was then forwarded for the Modeller analysis.

### Primary template and structure prediction using 3D-JIGSAW v.3.0

Upon input of the FAM72A sequence (149 AAs) into the 3D-JIGSAW v.3.0 server, the 3GA3\_A (133 AAs) template was suggested as the best homologous template for the 3D FAM72A protein structure (which is the same template suggestion as obtained with Phyre2) (Supplementary Fig. S8). Five structures were initially proposed, out of which the best

structure was selected based on data obtained, which were prechecked by various means such as Ramachandran plot and others (Supplementary Figs. S9–S13). It is noticeable that using the same template (3GA3\_A, having 133 AAs), Phyre2 predicted a 3D FAM72A protein structure containing 128 AAs. These 128 AAs cover the region from AA14 to AA141 of FAM72A (149 AAs). The 3DJIGSAW v.3.0 server also suggested 3GA3\_A (133 AAs) as a best template for FAM72A. In this predicted structure, however, the 3D FAM72A protein structure was composed of 133 AAs, which is the same number of AAs found in 3GA3\_A itself. These 133 AAs actually cover the entire 149 AAs of FAM72A with three gaps, including a total of 16 AAs. The selected template was then forwarded for the Modeller analysis.

#### Primary template and structure prediction using Swiss Model

Submission of FAM72A (149 AAs) to the Swiss Model server generated three 3D FAM72A protein structure models using three different templates (3MCA\_B (390 AAs), 1I8D\_A (213 AAs), and 1I8D\_B (213 AAs)). The proposed 3D FAM72A protein structures were prechecked by various means such as Ramachandran plot and others (Supplementary Figs. S14–20). The selected templates (3MCA\_B, 1I8D\_A, and 1I8D\_B) were then forwarded for the Modeller analysis.

#### Primary template and structure prediction using RaptorX

Submission of the FAM72A protein sequence (149 AAs) to the RaptorX server generated one full-length and two short versions of 3D FAM72A protein structure models (domains 1 and 2), using three different templates (4M0M\_A (756 AAs), 2FJA\_A (643 AAs), and 3UK7\_A (396 AAs)). The predicted full-length 3D FAM72A protein structure consisted of 145 AAs, and the respective template suggestion was 4M0M\_A (756 AAs). In the cases of the short-length versions, a total of ten 3D FAM72A protein structures were suggested (five structures for each domain (details are explained in the flow chart in the supplementary data file: Supplementary Figs. S23–S27)), out of which the two best short versions (2FJA\_A (643 AAs) and 3UK7\_A (396 AAs)) were selected. The two short-length versions (domains 1 and 2) were finally selected based on data obtained, which were prechecked by various means such as Ramachandran plot and others. All three of the identified templates (4M0M\_A, 2FJA\_A, and 3UK7\_A) were then forwarded for the Modeller analysis.

#### Primary template and structure prediction using multiple templates with I-TASSER

The I-TASSER server suggested ten templates to then generate five 3D FAM72A protein structure models, all of which contained 149 AAs and out of which the best one was selected

based on data obtained, which were prechecked by various means such as Ramachandran plot (Supplementary Figs. S28–S33).

#### Protein structure prediction with Modeller v.9.13

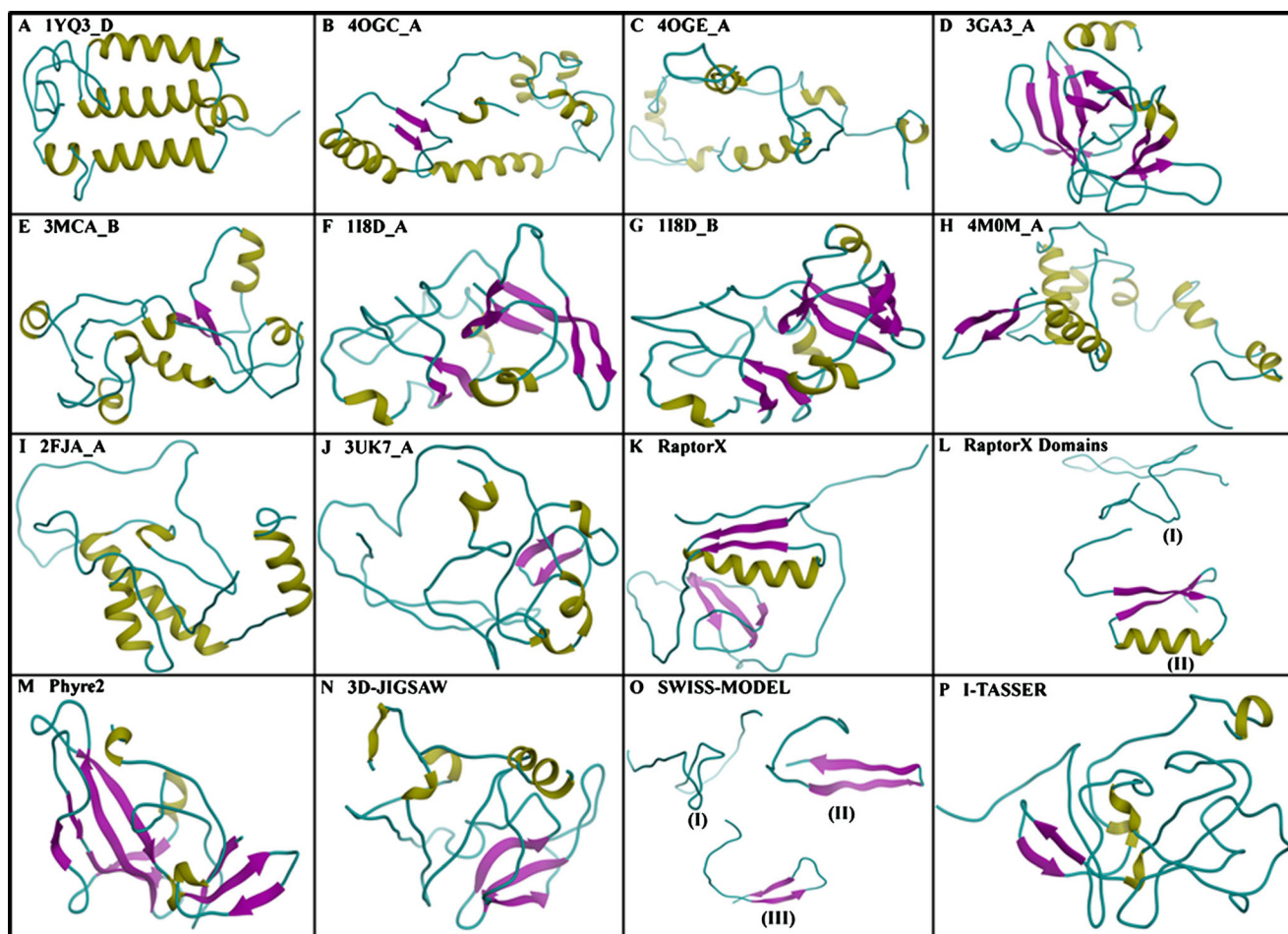
Next, we applied the Modeller (v.9.13) analysis [43–45] to further optimize the 3D FAM72A protein structure. Nine prechecked templates (1YQ3\_D, 4OGC\_A, 4OGE\_A, 3GA3\_A, 3MCA\_B, 1I8D\_B, 4M0M\_A, 2FJA\_A, and 3UK7\_A (obtained from NCBI-PDB-, Phyre2-, 3D-JIGSAW-, Swiss Model-, and RaptorX-based FAM72A models) were finally entered into the Modeller software analysis program, and data obtained were validated by various means, such as Ramachandran plot and others (Supplementary Figs. S34–S53). Furthermore, the obtained sequences of the structures (149 AAs) were always reversely confirmed by creating the sequence from the PDB file at <http://swift.cmbi.ru.nl/servers/html/soupir.html>. For further evaluation of data obtained, all of the structures were cross-aligned to each other and evaluated based on TM-scores normalized by the target length (Supplementary Figs. S54–S119) [12, 14, 46].

Thus, taken together, the findings of our *in silico* 3D FAM72A protein structure, finally modeled with Modeller 9.13 based on a variety of pre-predicted, preselected, and pre-evaluated templates (NCBI's PDB, Phyre2, 3DJIGSAW v.3.0, Swiss Model, RaptorX, and I-TASSER), are comprehensively presented in Fig. 1. Predicted structures are presented as the objective of Critical Assessment of Protein Structure Prediction (CASP; <http://predictioncenter.org/>) [47, 48] to assesses the ability of the various applied predictors (methods) to model protein structures in two different ways: (1) template-based modeling of the 3D FAM72A protein structure (Fig. 1a (1YQ3\_D) –o (Swiss Model); here, the objective of modeling is to predict the 3D FAM72A protein structures with an identified AA sequence by a similarity search of experimentally obtained 3D protein structures and followed by modeling (based on the related structures as templates)) and (2) free modeling (Fig. 1p (I-TASSER); here, objective of modeling is to predict the 3D FAM72A protein structures with multiple templates/threading using *de novo/ab initio* method).

#### Protein structure evaluation

The summarized stereochemical qualities of the obtained 3D FAM72A protein structure models are summarized in Table 1 (detailed explanations are provided in the supplementary data file: Supplementary Fig. S6–S53). The 3D FAM72A protein structure, based on the 3GA3\_A template (which was originally suggested by Phyre2 and 3DJIGSAW) and modeled with Modeller 9.13, appeared to be the best and most reliable structure in terms of stereochemical properties and overall





**Fig. 1** Assorted 3D FAM72A protein structures, predicted by a variety of methods, are shown. The 3D FAM72A protein structures (a–j) were modeled with Modeller 9.13 using the various templates suggested by searches with NCBI-PDB (a 1YQ3\_D, b 4OGC\_A, and c 4OGE\_A), Phyre2 (d 3GA3\_A), 3D-JIGSAW v.3.0 (d 3GA3\_A), Swiss Model (e 3MCA\_A, f 1I8D\_A, and g 1I8D\_B), and RaptorX (h 4M0M\_A, i 2FJA\_A, and j 3UK7\_A). k Suggested 3D FAM72A protein structure predicted by RaptorX with 4M0M\_A as a template. l Suggested 3D FAM72A protein structures for two short domains from RaptorX: (I) range, 76 to 134 (59-amino acids), based on the 2FJA\_A template, and (II) range, 1 to 51 (51-amino acids), based on the 3UK7\_A template. m

Suggested (Phyre2) 3D FAM72A protein structure having 128 AAs based on the 3GA3\_A template. n Suggested (3D-JIGSAW v.3.0) 3D FAM72A protein structure having 133 AAs, based on the 3GA3\_A template. o Three suggested (Swiss Model) 3D FAM72A protein structures: (I) range, 97 to 137 (41-amino acids), based on the 3MCA\_B template, (II) range, 97 to 129 (33-amino acids), based on the 1I8D\_A template, and (III) range, 97 to 129 (33-amino acids), based on 1I8D\_B template are shown. p Suggested (I-TASSER) 3D FAM72A protein structure based on multiple templates. Legend: (i) yellow, helix; (ii) purple,  $\beta$ -sheet strand; and (iii) cyan, coil structure

geometry compared with the other structures (Figs. 1d and 2a). We observed that 136 AAs of the FAM72A sequence are non-glycine and non-proline AA residues (149 AAs; 2 AAs are end-residues (excluding Gly and Pro), 7 AAs are Gly residues, and 4 AAs are Pro residues) out of which 120 AAs (88.2 % of 136 AAs) belonged to the most favored regions (A, B, L) in the Ramachandran plot. This 3D protein structure also has a significant overall  $G$ -factor value of  $-0.2$ . Ideally, a  $G$ -factor should be above  $-0.5$  [49–52], which provides a measure of normal stereochemical properties and overall geometry of the predicted 3D FAM72A protein structure (Supplementary Fig. S41).

#### Protein structure comparison analysis by alignment

We applied the TM-align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>), which is a highly optimized algorithm for 3D protein structure comparison and alignment, to evaluate the topological/geometrical similarity between structures. The 3D structural relationships among the various modeled 3D FAM72A protein structures allowed us to investigate correlations in topological similarity of the theoretical developed 3D protein structures, which were normalized to the target length of 149 AAs (Supplementary Figs. S54–S125). The aligned length (no. of AA residues) and comprehensively analyzed

**Table 1** Summary of stereochemical validation parameters of the predicted 3D FAM72A protein structures

| Template                    | Procedure                       | No. of AAs in suggested template | Similar AAs (identical and positive) with FAM72A | No. of AAs in predicted or suggested structure | PROCHECK validation parameters   |                                       |
|-----------------------------|---------------------------------|----------------------------------|--|--|--|---------------------------------------|
|                             |                                 |                                  |  |  | Residues in most favored regions (A, B, L) in Ramachandran plot <sup>a</sup> (%) | Overall <i>G</i> -factor <sup>b</sup> |
| 1YQ3_D                      | 1. NCBI-PDB                     | 103                              | 19   | –  | –  | –                                     |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 85.3   | –0.1                                  |
| 4OGC_A                      | 1. NCBI-PDB                     | 1,101                            | 15   | –  | –  | –                                     |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 86.8   | –0.2                                  |
| 4OGE_A                      | 1. NCBI-PDB                     | 1,101                            | 15   | –  | –  | –                                     |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 85.3   | –0.2                                  |
| 3GA3_A                      | 1. Phyre2                       | 133                              | 19   | 128  | 85.0   | –0.4                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 88.2   | –0.2                                  |
| 3GA3_A                      | 1. 3D-JIGSAW                    | 133                              | 19   | 133  | 66.7   | –0.3                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 88.2   | –0.2                                  |
| 3MCA_B                      | 1. Swiss Model                  | 390                              | 11   | 41   | 71.4   | –0.3                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 80.1   | –0.2                                  |
| 1I8D_A                      | 1. Swiss Model                  | 213                              | 5  | 33   | 75.0   | –0.2                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 84.6   | –0.3                                  |
| 1I8D_B                      | 1. Swiss Model                  | 213                              | 5  | 33   | 78.6   | –0.5                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 81.6   | –0.2                                  |
| 4M0M_A                      | 1. RaptorX                      | 756                              | 8  | 145  | 86.4   | –0.3                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 81.6   | –0.2                                  |
| 2FJA_A                      | 1. RaptorX                      | 643                              | –  | 59   | 86.0   | –0.1                                  |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 81.6   | –0.2                                  |
| 3UK7_A                      | 1. RaptorX                      | 396                              | –  | 51   | 91.3   | 0.0                                   |
|                             | 2. Forwarded to Modeller        |                                  |  | 149  | 83.8   | –0.2                                  |
| Multiple template threading | 1. I-TASSER developed structure | –                                | –  | 149  | 72.1   | –0.7                                  |

<sup>a</sup> Ideally, a good quality model would be expected to have over 90 % of the residues in these “most favorable/core” regions. The percentage of residues in the “most favorable/core” regions is one of the better guides to stereochemical quality

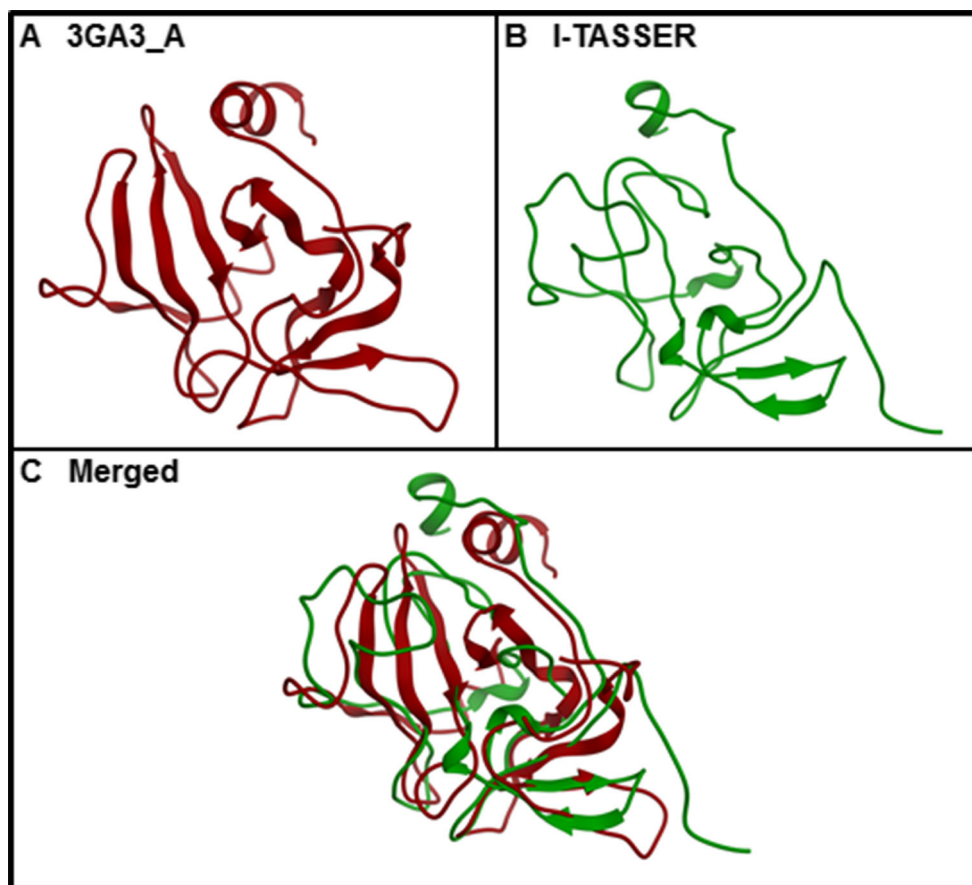
<sup>b</sup> Ideally, *G*-factor should be above –0.5. Values below –1.0 may need investigation. The *G*-factor measures overall geometry of a 3D protein structure. The *G*-factor provides a measure of how “normal,” or alternatively how “unusual,” a given stereochemical property is. In PROCHECK, it is computed for the following properties (torsion angles: (i) phi-psi combination, (ii) chi1-chi2 combination, (iii) chi1 torsion for those residues that do not have a chi-2, (iv) combined chi-3 and chi-4 torsion angles, and (v) omega torsion angles; covalent geometry: (i) main-chain bond lengths and (ii) main-chain bond angles). When applied to a given residue, a low *G*-factor indicates that the property corresponds to a low-probability conformation. So, for example, residues falling in the disallowed regions of the Ramachandran plot will have a low (or very negative) *G*-factor and similarly for unfavorable chi1-chi2 and chi1 values also. Thus, if a protein has many residues with low *G*-factors, it suggests that something may be amiss with its overall geometry ([http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=n/a&template=doc\\_procheckgf.html](http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=n/a&template=doc_procheckgf.html)).

statistical validation parameters (TM-score and RMSD) are explained in Supplementary Table S1. Interestingly, our results show that the 3D FAM72A protein structure model based on the 3GA3\_A template, which was originally suggested by Phyre2 and 3DJIGSAW and finally modeled with Modeller 9.13 (Fig. 2a), and the ab initio-modeled 3D FAM72A protein structure based on multiple templates from I-TASSER (Fig. 2b), have significant geometrical structural similarity (Fig. 2c). The alignment of these two modeled structures shows that an aligned length covering 121 AA residues has an RMSD value of 3.04. The TM-scores of 0.62838, if normalized by the length of the

structure for 3GA3\_A (Fig. 2a), and 0.62838, if normalized by the length of the structure for I-TASSER (Fig. 2b), demonstrate the natural stereochemical attributes of the predicted 3D FAM72A protein structure.

Furthermore, the final refined 3D FAM72A protein structure model (based on the 3GA3\_A template, which was originally suggested by Phyre2 and 3DJIGSAW and modeled with Modeller 9.13) was also superimposed with the structure of the original template 3GA3\_A itself to countercheck in silico the 3D FAM72A protein structure’s originality compared with that of its template (3GA3\_A) by using TM-align (Supplemental Fig. S124).

**Fig. 2** 3D FAM72A protein structure alignment. **a** 3D FAM72A protein structure based on the 3GA3\_A template and modeled with Modeller 9.13. **b** 3D FAM72A protein structure from I-TASSER. **c** 3D FAM72A protein structure alignment of structures shown in (a) and (b). The alignment shows that an aligned length covering 121-amino acid residues has an RMSD value of 3.04. The TM-score is 0.62838 (if normalized by the length of the structure for 3GA3\_A (a)) or 0.62838 (if normalized by the length of the structure for I-TASSER (b))



In addition to these 3D structure alignments, sequence alignments with different matrices (PAM-30, PAM-70, PAM2-50, BLOSUM-80, BLOSUM-62, BLOSUM-45, BLOSUM-50, and BLOSUM-90) for the various suggested templates (NCBI-PDB database, Phyre2, 3DJIGSAW, and Swiss Model) for the 3D FAM72A protein structure showed comparative similarities among the various templates (Supplementary Figs. S126–S127). Consequently, a conclusive inference can be drawn (Table 1; Fig. 2; Supplemental Figs. S124–125; Supplemental Table S1) that the 3D FAM72A protein structure based on the 3GA3\_A template (which was originally suggested by Phyre2 and 3DJIGSAW and finally modeled with Modeller 9.13) is the most reliable predicted 3D FAM72A protein structure with respect to model-model and model-template correlation. Therefore, this theoretical 3D FAM72A protein structure was chosen for further ligand-protein interaction prediction analyses.

#### Lead discovery by protein-ligand-binding-site prediction

The COACH, TM-SITE, S-SITE, COFACTOR, and ConCavity approaches (e.g., via <http://zhanglab.cmb.med.umich.edu/>) suggested potential ligand-binding sites of the FAM72A protein with various molecules based on a BioLiP database screening (<http://zhanglab.cmb.med.umich.edu/>

[BioLiP/index.html](http://zhanglab.cmb.med.umich.edu/BioLiP/index.html)). The predicted results indicate that FAM72A can interact with  $Zn^{2+}$  (Fig. 3a) and  $Fe^{3+}$  ions, nucleic acids, and the organic compound RSM: (2s)-2-(acetylamino)-N-methyl-4-[(R)-methylsulfinyl] butanamide. A detailed explanation is provided in Supplementary Figs. S128–S143. The basic structure (Fig. 3b) contains two H-bond donor groups and four H-bond acceptors (Supplemental Fig. S142). Additionally, pharmacophore analysis was performed using the PharmMapper server to detect the basic pharmacophore group of RSM molecule [53]. Active molecular pharmacophore features H-bond donor groups (lime green) and H-bond acceptors (magenta) have been defined (Fig. 4). The formation of a hydrogen bond is pivotal to defining the 3D molecular structure and function as well as the formation of ligand-protein complexes. Moreover, the affinity of ligands to form hydrogen bonds is an important issue for any ligand design. Of course, hydrogen bonding also affects the delivery and distribution of drugs within the biological system [54].

The levels of reliability of the various ligand-binding sites were checked based on their confidence score (C-score) statistical validation parameters. The COACH (C-score=0.30), TM-SITE (C-score=0.38), and COFACTOR (C-score=0.37) approaches all suggested the same, mostly favorable  $Zn^{2+}$ -binding sites, including Cys18, Cys21, Cys74, and Cys77



(Fig. 3a). Additionally, according to the TM-SITE approach, the  $\text{Fe}^{3+}$  ion can also bind with these same Cys-binding sites. In addition to the ion-binding sites, the RSM molecule-binding sites can be considered an important FAM72A interaction because this interaction could be used as an unconventional preliminary stage for potential antitumorigenic drug screening tests. The COFACTOR approach suggested that the RSM molecule has a binding capacity for the FAM72A protein (C-score=0.02; Supplementary Fig. S141), and the RSM-binding sites identified on the FAM72A protein are Tyr83, Val85, Cys96, and Asn97 (Fig. 3b). Subsequently, the binding site prediction with shape-based ligand matching with binding pocket (BSP-SLIM) molecular docking method [55] was applied to accurately redefine the binding site of RSM with FAM72A. We observed that potential RSM-binding sites, based on the COFACTOR analysis approach, are Tyr83, Val85, Cys96, and Asn97 (there are other AAs such as Cys67, Leu69, Lys63, and Glu39 which may also interact with FAM72A (Fig. 3b)), and we also found that BSP-SLIM molecular docking suggested RSM-binding sites (docking score=4.979) are Tyr83, Val85, Glu39, Leu69, and Lys63 (Supplementary Fig. S144). Therefore, both binding-site predictions (based on COFACTOR and BSP-SLIM molecular docking) suggested partially the same binding sites (e.g., Tyr83 and Val85) of RSM with FAM72A (Supplementary Fig. S144; Fig. 3b).

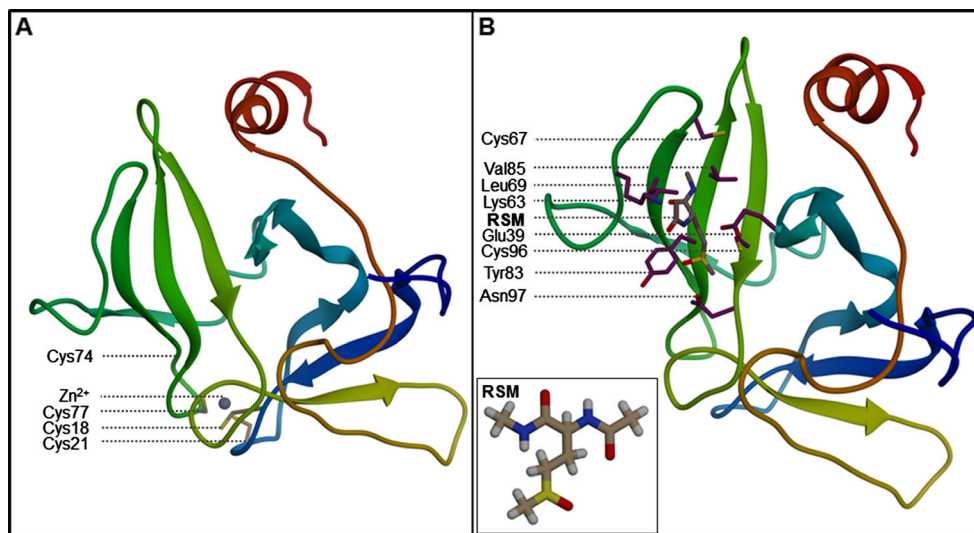
## Discussion

The advances in biocomputational analyses have opened new options for the development of novel approaches for a 3D protein structure analysis [12–16]. In our present study on the 3D FAM72A protein structure modeling using multiple

approaches, we found that the template 3GA3\_A is the best model for the 3D FAM72A protein structure prediction. Structural topological/geometrical similarity analyses between the 3GA3\_A-template-based model of the 3D FAM72A protein structure (template 3GA3\_A modeled with Modeller 9.13) and the ab initio-modeled structure of the 3D FAM72A protein structure (using I-TASSER) showed a significant correlation between the two structures in terms of secondary structure ( $\alpha$ -helix,  $\beta$ -sheet, and coil) and tertiary structure (protein folding) (Fig. 2). The structure evaluation (Supplementary Figs. S6–S53) and 3D alignment (Supplementary Figs. S55–S125) show that the 3D FAM72A protein structure based on the 3GA3\_A template (originally suggested by Phyre2 and 3DJIGSAW analyses and subsequently modeled with Modeller 9.13) can be considered as the most reliable 3D protein structure for FAM72A compared with the other models suggested. In considering the overall *G*-factor, which defines the stereochemical quality of the 3D protein structure, the full length 3D FAM72A protein structure based on 1YQ3\_D and modeled with Modeller 9.13 (originally suggested by NCBI-PDB database as one of the best possible templates for the FAM72A structure (Supplementary Fig. S3)) has the best overall *G*-factor, with a value of  $-0.1$  (Table 1). However, the model-template 3D alignment along with the model-model 3D alignment analyses and evaluations (Supplementary Table S1) all reveal that the modeled 3D FAM72A protein structure based on 3GA3\_A (Fig. 3) is the most reliable model and is also better than the model based on the 1YQ3\_D template for the 3D FAM72A protein structure (Supplementary Fig. S125).

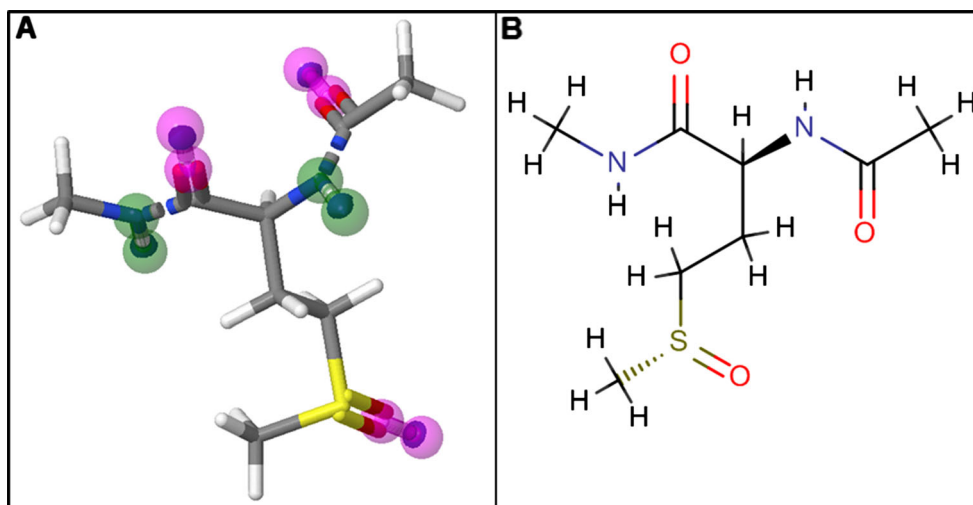
In the context of CASP's assessment [56–58] for the advancement of methods of identifying 3D protein structures from their AA sequences, we observed that the template-based (3GA3\_A) 3D FAM72A protein structure (with an overall *G*-factor= $-0.2$ ) has a much reliable overall stereochemical and

**Fig. 3** FAM72A– $\text{Zn}^{2+}$  and RSM interactions. **a**  $\text{Zn}^{2+}$ -binding sites are shown on the developed 3D FAM72A protein structure. Suggested  $\text{Zn}^{2+}$ -binding sites on the FAM72A protein are Cys18, Cys21, Cys74, and Cys77. **b** The RSM ((2*s*)-2-(acetylamino)-*N*-methyl-4-[(*R*)-methylsulfinyl]butanamide)-binding sites are shown on the developed 3D FAM72A protein structure. Suggested RSM-binding sites on FAM72A protein are Tyr83, Val85, Cys96, and Asn97





**Fig. 4 a** The basic pharmacophore groups of RSM ((2*s*)-2-(acetylamino)-*N*-methyl-4-[(*R*)-methylsulfinyl]butanamide) are shown: H-bond donor groups (lime green) and H-bond acceptors (magenta). Atomic color legend: (i) blue, nitrogen; (ii) yellow, sulfur; (iii) red, oxygen; (iv) gray, carbon, and (v) colorless, hydrogen. **b** The basic atomic structure of RSM (molecular formula, C<sub>8</sub>H<sub>16</sub>N<sub>2</sub>O<sub>3</sub>S) is shown



geometrical quality compared with the de novo/ab initio (I-TASSER) 3D FAM72A protein structure prediction (overall *G*-factor=−0.7).

Binding interactions of FAM72A with Zn<sup>2+</sup>, Fe<sup>3+</sup>, and nucleic acid indicate that this protein may have a significant cellular function as a transcription factor, a finding that might be particularly interesting in view of FAM72A's role in cancer cell signaling [3] and may also interfere with enzymatic redox reactions [59–61].

Additionally, the physicochemical interactions of RSM with the FAM72A protein may be a valuable asset in developing screening tests for antitumorigenic therapeutics. The RSM molecule is also a bound ligand of methionine sulfoxide reductases (3HCl\_A having 154 AAs, 3HCH\_A having 145 AAs, and 3HCH\_B having 146 AAs), which reduce methionine sulfoxide to methionine [62]. The molecular features of the RSM molecule (e.g., H-bond donor groups and H-bond acceptor groups) suggest the possibility that RSM can interact with FAM72A via formation of hydrogen bonds to exert a possible anti-FAM72A activity. Despite significant progress in de novo design of ligands, getting a preliminary basic molecule is the decisive step in any ligand-target design. The *FAM72A* gene has four exons, and the translated exons are indicated in the predicted 3D FAM72A protein structures (Supplementary Fig. S145). The suggested RSM-binding sites are also indicated therein with Tyr83, Val85, Cys96, and Asn97 on the  $\beta$ -sheet and AA78 to AA119 included, all of which are translated from *FAM72A* gene's exon 3 (Supplementary Fig. S145d; Fig. 3). Present finding based on structure-based drug design suggests that the lead molecule RSM can be used for the future setup of high-throughput screening (HTS) experiments with RSM as seed molecule and also to understand the possible biochemical anti-FAM72A interaction-activity between RSM and FAM72A in a biological system [63, 64]. The identification and characterization of cancer predisposing genes as well as the

development of cancer-type- and gene-specific therapeutic drugs remains a major challenge in the pharmaceutical industries [65–69]. Because of this, our in silico 3D structure modeling of the novel FAM72A protein and its associated ligand-protein, as described here, and protein-protein [3] interactions may provide valuable information for future experimental in vivo cancer research.

## Conclusions

Cancer patients still lack sufficient treatments due to the deficiency of targeted therapeutics. In the context of genome instability and genetic diversity between cancer cells and normal cells, targeted therapies against a gene-specific target based on its biological protein function in cancer is an urgent need rather than nontargeted therapies without a priori knowledge of the targets. The most fundamental aspect of a protein molecule is its geometrical structure (folding). Therefore, we aimed to develop proof of principle theoretical strategies to demonstrate the potential utility of this novel approach for both the development of a 3D FAM72A protein structure and the determination of unique specific novel targets for FAM72A for the treatment of various types of cancer [3, 4]. Our data provide the first starting point for contemporary in silico 3D structure modeling, uncovering FAM72A protein-ligand-binding sites, which can be further investigated by in vitro and in vivo experiments to confirm the effectiveness of the suggested compound as lead for screening tests in search of novel drugs for the treatment of FAM72A-based cancers. We hope these results may assist multidimensional cancer research in controlling various cancer types that are of clinical and societal importance across the globe.

**Acknowledgments** This study was supported by Hanyang University. The authors declare no competing interests.

## References

- Jung KW, Won YJ, Kong HJ, Oh CM, Seo HG, Lee JS. Cancer statistics in Korea: incidence, mortality, survival and prevalence in 2010. *Cancer Res Treat.* 2013;45:1–14.
- Nehar S, Mishra M, Heese K. Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *FEBS Lett.* 2009;583:3247–53.
- Heese K. The protein p17 signaling pathways in cancer. *Tumour Biol.* 2013;34:4081–7.
- Guo C, Zhang X, Fink SP, Platzner P, Wilson K, Willson JK, et al. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res.* 2008;68:6118–26.
- Jorgensen WL. Drug discovery: pulled from a protein's embrace. *Nature.* 2010;466:42–3.
- Jorgensen WL. Foundations of biomolecular modeling. *Cell.* 2013;155:1199–202.
- Renault L, Chou HT, Chiu PL, Hill RM, Zeng X, Gipson B, et al. Milestones in electron crystallography. *J Comput Aided Mol Des.* 2006;20:519–27.
- Rusche B. The 3Rs and animal welfare—conflict or the way forward? *ALTEX.* 2003;20:63–76.
- Jorgensen WL. The many roles of computation in drug discovery. *Science.* 2004;303:1813–8.
- Michel J, Essex JW. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J Comput Aided Mol Des.* 2010;24:639–58.
- Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. *J Comput Aided Mol Des.* 2012;26:15–26.
- Peng J, Xu J. A multiple-template approach to protein threading. *Proteins.* 2011;79:1930–9.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5:725–38.
- Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX Web server. *Nat Protoc.* 2012;7:1511–22.
- Ma J, Wang S, Zhao F, Xu J. Protein threading using context-specific alignment potential. *Bioinformatics.* 2013;29:i257–65.
- Sanchez-Gonzalez G, Kim JK, Kim DS, Garduno-Juarez R. A beta-complex statistical four body contact potential combined with a hydrogen bond statistical potential recognizes the correct native structure from protein decoy sets. *Proteins.* 2013;81:1420–33.
- Coordinators NR. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2014;42:D7–17.
- Zhang Y. I-TASSER server for protein 3d structure prediction. *BMC Bioinformatics.* 2008;9:40.
- Webb B, Sali A. Protein structure modeling with modeller. *Methods Mol Biol.* 2014;1137:1–15.
- Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, et al. MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 2014;42:D336–46.
- Moreno-Vargas L, Correa-Basurto J, Maroun RC, Fernandez FJ. Homology modeling of the structure of acyl CoA:isopenicillin *n*-acyltransferase (IAT) from penicillium chrysogenum IAT interaction studies with isopenicillin-*n*, combining molecular dynamics simulations and docking. *J Mol Model.* 2012;18:1189–205.
- Jakubik J, Randakova A, Dolezal V. On homology modeling of the M(2) muscarinic acetylcholine receptor subtype. *J Comput Aided Mol Des.* 2013;27:525–38.
- Dhanavade MJ, Jalkute CB, Barage SH, Sonawane KD. Homology modeling, molecular docking and md simulation studies to investigate role of cysteine protease from *Xanthomonas campestris* in degradation of amyloid beta peptide. *Comput Biol Med.* 2013;43:2063–70.
- Nicolau Jr N, Giuliani S. Modeling and molecular dynamics of the intrinsically disordered e7 proteins from high- and low-risk types of human papillomavirus. *J Mol Model.* 2013;19:4025–37.
- Tarhda Z, Semlali O, Kettani A, Moussa A, Abumrad NA, Ibrahim A. Three dimensional structure prediction of fatty acid binding site on human transmembrane receptor CD36. *Bioinform Biol Insights.* 2013;7:369–73.
- Yadav BS, Tripathi V, Kumar A, Khan MF, Barate A, Kumar A, et al. Molecular modeling and docking characterization of Dectin-1 (PAMP) receptor of *Bubalus bubalis*. *Exp Mol Pathol.* 2012;92:7–12.
- Basta HA, Sgro JY, Palmenberg AC. Modeling of the human rhinovirus C capsid suggests a novel topography with insights on receptor preference and immunogenicity. *Virology.* 2014;448:176–84.
- Moreira LG, Pereira LC, Drummond PR, De Mesquita JF. Structural and functional analysis of human SOD1 in amyotrophic lateral sclerosis. *PLoS ONE.* 2013;8:e81979.
- Ashok Kumar HG, Venkatesh YP. In silico analyses of structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic thaumatin-like protein in comparison with allergenic plant TLPs. *Mol Immunol.* 2014;57:119–28.
- Lodhi SS, Farmer R, Singh AK, Jaiswal YK, Wadhwa G. 3d structure generation, virtual screening and docking of human Ras-associated binding (Rab3a) protein involved in tumorigenesis. *Mol Biol Rep.* 2014;41:3951–9.
- Dhanavade MJ, Sonawane KD. Insights into the molecular interactions between aminopeptidase and amyloid beta peptide using molecular modeling techniques. *Amino Acids.* 2014
- Chaturvedi P, Dhanik M, Pande A. Characterization and structural analysis of hepcidin like antimicrobial peptide from *Schizothorax richardsonii* (Gray). *Protein J.* 2014;33:1–10.
- de Carvalho MD, De Mesquita JF. Structural modeling and in silico analysis of human superoxide dismutase 2. *PLoS ONE.* 2013;8:e65558.
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR.* 1996;8:477–86.
- Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005;33:2302–9.
- Hadley C, Jones DT. A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. *Structure.* 1999;7:1099–112.
- Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr.* 2010;66:133–44.
- Lomb L, Barends TR, Kassemeyer S, Aquila A, Epp SW, Erk B, et al. Radiation damage in protein serial femtosecond crystallography using an X-ray free-electron laser. *Phys Rev B Condens Matter Mater Phys.* 2011;84:214111.
- Betancourt MR, Skolnick J. Universal similarity measure for comparing protein structures. *Biopolymers.* 2001;59:305–9.
- Yang J, Roy A, Zhang Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* 2013;41:D1096–103.
- Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics.* 2013;29:2588–95.
- Ruan P, Hayashida M, Maruyama O, Akutsu T. Prediction of heterotrimeric protein complexes by two-phase learning using neighboring kernels. *BMC Bioinformatics.* 2014;15(2):S6.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics.* 2006;5:5–6.
- Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* 2000;29:291–325.

45. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234:779–815.
46. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins.* 2004;57:702–10.
47. Tai CH, Bai H, Taylor TJ, Lee B. Assessment of template-free modeling in CASP10 and ROLL. *Proteins.* 2014;82 Suppl 2:57–83.
48. Kryshchak A, Monastyrskyy B, Fidelis K. CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. *Proteins.* 2014;82 Suppl 2:7–13.
49. Baker BR, Garrell RL. G-factor analysis of protein secondary structure in solutions and thin films. *Faraday Discuss.* 2004;126:209–22. discussion 45–54.
50. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins.* 1992;12:345–64.
51. Groll M, Bochtler M, Brandstetter H, Clausen T, Huber R. Molecular machines for protein degradation. *ChemBiochem.* 2005;6:222–56.
52. McPhie P. Circular dichroism studies on proteins in films and in solution: estimation of secondary structure by G-factor analysis. *Anal Biochem.* 2001;293:109–19.
53. Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, et al. PharmMapper server: a Web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* 2010;38:W609–14.
54. Kubinyi H. Drug research: myths, hype and reality. *Nat Rev Drug Discov.* 2003;2:665–8.
55. Lee HS, Zhang Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins.* 2012;80:93–110.
56. Karakas M, Woetzel N, Meiler J. BCL::contact-low confidence fold recognition hits boost protein contact prediction and de novo structure determination. *J Comput Biol.* 2010;17:153–68.
57. Kryshchak A, Moulton J, Bales P, Bazan JF, Biasini M, Burgin A, et al. Challenging the state of the art in protein structure prediction: highlights of experimental target structures for the 10th critical assessment of techniques for protein structure prediction experiment CASP10. *Proteins.* 2014;82 Suppl 2:26–42.
58. Moulton J, Fidelis K, Kryshchak A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins.* 2014;82 Suppl 2:1–6.
59. Klug A. Zinc finger peptides for the regulation of gene expression. *J Mol Biol.* 1999;293:215–8.
60. Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci.* 2007;32:63–70.
61. Zhong F, Lisi GP, Collins DP, Dawson JH, Pletneva EV. Redox-dependent stability, protonation, and reactivity of cysteine-bound heme proteins. *Proc Natl Acad Sci U S A.* 2014;111:E306–15.
62. Ranaivosoa FM, Neiers F, Kauffmann B, Boschi-Muller S, Branlant G, Favier F. Methionine sulfoxide reductase B displays a high level of flexibility. *J Mol Biol.* 2009;394:83–93.
63. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov.* 2011;10:188–95.
64. Keseru GM, Makara GM. The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov.* 2009;8:203–12.
65. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646–74.
66. Hainaut P, Plymth A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Curr Opin Oncol.* 2013;25:50–1.
67. Rahman N. Realizing the promise of cancer predisposition genes. *Nature.* 2014;505:302–8.
68. Zheng YW, Tsuchida T, Taniguchi H. A novel concept of identifying precancerous cells to enhance anti-cancer therapies. *J Hepatobiliary Pancreat Sci.* 2012;19:621–5.
69. Zieba A, Grannas K, Soderberg O, Gullberg M, Nilsson M, Landegren U. Molecular tools for companion diagnostics. *N Biotechnol.* 2012;29:634–40.