Briefings in Bioinformatics, 00(00), 2021, 1–11

doi: 10.1093/bib/bbaa373
Problem Solving Protocol

HISNAPI: a bioinformatic tool for dynamic hot spot analysis in nucleic acid–protein interface with a case study

Long-Can Mei[†], Yu-Liang Wang[†], Feng-Xu Wu, Fan Wang, Ge-Fei Hao and Guang-Fu Yang

Corresponding author: Ge-Fei Hao, Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, China. State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Research and Development Center for Fine Chemicals, Guizhou University, Guiyang 550025, China. Tel.: +86-27-67867706; Fax: +86-27-67867706; E-mail: gefei_hao@foxmail.com.
[†]These authors contributed equally to this work.

Abstract

Protein–nucleic acid interactions play essential roles in many biological processes, such as transcription, replication and translation. In protein–nucleic acid interfaces, hotspot residues contribute the majority of binding affinity toward molecular recognition. Hotspot residues are commonly regarded as potential binding sites for compound molecules in drug design projects. The dynamic property is a considerable factor that affects the binding of ligands. Computational approaches have been developed to expedite the prediction of hotspot residues on protein–nucleic acid interfaces. However, existing approaches overlook hotspot dynamics, despite their essential role in protein function. Here, we report a web server named Hotspots *In silico* Scanning on Nucleic Acid and Protein Interface (HISNAPI) to analyze hotspot residue dynamics by integrating molecular dynamics simulation and one-step free energy perturbation. HISNAPI is capable of not only predicting the hotspot residues in protein–nucleic acid interfaces but also providing insights into their intensity and correlation of dynamic motion. Protein dynamics have been recognized as a vital factor that has an effect on the interaction specificity and affinity of the binding partners. We applied HISNAPI to the case of SARS-CoV-2 RNA-dependent RNA polymerase, a vital target of the antiviral drug for the treatment of coronavirus disease 2019. We identified the hotspot residues and characterized their dynamic behaviors, which might provide insight into the target site for antiviral drug design. The web server is freely available via a user-friendly web interface at http://chemyang.ccnu.edu.cn/ccb/server/HISNAPI/ and http://agroda.gzu.edu.cn:9999/ccb/server/HISNAPI/.

Key words: protein–nucleic acid interaction; hotspot residues; alanine scanning; dynamic behavior; drug target sites

Introduction

Proteins and nucleic acids are the two fundamental and significant components of living organisms. Protein is the product of gene expression, and gene expression depends on protein. Protein–nucleic acid interactions exist at almost all levels of gene expression [1, 2]. In protein–nucleic acid interfaces, a residue whose mutation to alanine leads to a large reduction in the binding free energy is termed as hotspot [3–5]. They are often clustered and packed to form 'hot regions' [6]. Typically, the technique of per-residue binding free energy decomposition

Long-Can Mei is a PhD student at College of Chemistry, Central China Normal University (CCNU). The direction of his thesis is biomolecular simulation.
Yu-Liang Wang is a master student at College of Chemistry of CCNU, the direction of his thesis is quantitative structure–activity relationship.
Feng-Xu Wu is a PhD student at College of Chemistry of CCNU, the direction of his thesis is rational drug design.
Fan Wang is an associate professor at College of Chemistry of CCNU. He received the PhD degree in Computational Chemistry from University of Amiens, France.
Ge-Fei Hao is a professor in bioinformatics in College of Chemistry of CCNU. He received his PhD in Pesticide Science from CCNU.
Guang-Fu Yang is a professor in chemical biology. He is the group leader and received the PhD degree in Pesticide Science from Nankai University, Tianjin, China.

Submitted: 24 September 2020; **Received (in revised form):** 19 November 2020

is applied to identify hotspot residues [7–9]. Rigorous approaches have been developed to calculate the binding free energy, such as free energy perturbation (FEP) [10], thermodynamic integration (TI) [11] and Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA) [12]. The decomposition of free energy is performed toward individual residues. These hotspot residues are the main favorable contributors to the binding interactions between protein and nucleic acid molecules. Modulating these regions contributes to an increased knowledge of protein function and thus uncovers their pathological implications. For example, targeting the hot regions of the protein–nucleic acid interface is an important modality in the treatment of cancer [13–15]. Hence, studying favored nucleic acid binding hotspots on proteins may provide essential information for identifying the targetable region of the protein–nucleic acid interface in disease treatment.

Recently, the dynamic properties of hotspot are receiving more and more attention [16]. The flexibility of hotspots plays a key role in allosteric regulation of protein surfaces. The nucleic acid could induce conformational changes in hotspot residues to form a substantially concave topology, which may be a potential binding pocket in drug discovery [17]. Conformational change is important for the formation of binding pockets and interactions with other partners. Structural flexibility has an effect on conformational movement, which may change an existent pocket or form a new pocket. Hotspots *In silico* Scanning on Nucleic acid and Protein Interface (HISNAPI) gives a good description of the pattern of hotspot movement, which allows the regulation of pocket formation on binding campaigns. The dynamic properties of these binding pockets are crucial for drug binding affinity and specificity [18]. For example, the binding and dissociation constants vary with different dynamics in the case of a maltose-binding protein and a series of mutants [16]. Another example is the consideration of protein binding pocket dynamics in p38 mitogen-activated protein kinase, which helped to find an inhibitor [19]. Hence, knowledge of nucleic acid binding-induced protein hotspot dynamics may facilitate a more comprehensive understanding of protein function, which is particularly relevant to drug discovery.

To date, numerous computational methods based on various principles have been developed to accelerate the prediction of the effects of mutations on protein–nucleic acid binding. Scoring methods have been developed for evaluating protein–RNA binding affinity, such as QUASI-RNP and DARS-RNP [20], ITScore-PR [21] and 3dRPC-Score [22]. Machine learning-based methods have increasingly emerged to model binding interactions [23–27]. These approaches could be utilized in the form of standalone programs, while their operational complexity may daunt a researcher without skillful computational chemistry knowledge. Several methods have been developed into web servers for public use, such as mCSM-NA [28], SAMPDI [29], PrabHot [30] and PrPDH [31]. However, the goal of these methods is only to identify the hotspot residues, regardless of their dynamic properties associated with protein functions.

To this end, we introduced a web server, named HISNAPI, to assist in describing the dynamic motions of hotspot residues in protein–nucleic acid binding interactions. Molecular dynamic (MD) simulation and one-step FEP have been integrated to develop a fast and accurate *de novo* method for binding free energy calculations [32]. HISNAPI was validated using different types of cross-validation on diverse and large sets of single alanine mutations from the ProNIT [33] and dbAMEPNI [34] databases, and compared with several other tools. The correlation we achieved for the binding free energy changes of 299 mutants from 40 protein–nucleic acid systems was $R = 0.77$. The sensitivity, specificity and precision were 74.4, 87.8 and 81.6%, respectively, with AUC = 0.86. Because of conformational sampling by molecular simulations, HISNAPI has an exclusive advantage in generating the dynamic information contained in simulation trajectories. HISNAPI provides a reliable way to predict the effect of single alanine mutations on protein–DNA/RNA binding and to characterize the dynamic properties of hotspot amino acids.

Materials and methods

Interfacial residues detection

The interfacial residues are identified against a distance criterion: those amino acids within a distance cut-off from any atoms in a nucleic acid molecule are defined as interfacial residues. A default 4 Å cut-off is recommended, which is accepted in most cases [35].

MD simulation

Considering conformational flexibility, we performed conventional MD simulation on protein–nucleic acid complexes to obtain conformational ensembles by AMBER 16 program [36]. Each complex was prepared by the *tleap* module in AMBER. Protein and nucleic acid molecules were parameterized by ff14SB force field [37], RNA by RNA OL3 force field [38] and DNA by DNA OL15 force field [39]. Each system was solvated in a TIP3P water [40] box with 10 Å distance between the solute and box. Counter-ions, $Na+$ and $Cl-$, were added to neutralize the unbalanced charges. Following the standard protocol of energy minimization, each system was then minimized by two steps: firstly, all the heavy atoms in the backbone of the protein were restrained with an elastic constant of 50 $kcal \cdot mol^{-1} \cdot Å^{-2}$ (2500 cycles of steepest descent and 2500 cycles of conjugate gradient minimizations); secondly, the whole system was minimized for 5000 steps without any restraint (2500 cycles of steepest descent and 2500 cycles of conjugate gradient minimizations). Each system was gradually heated from 0 to 300 K during a period of 500 ps in the NVT ensemble. Then, 1 ns equilibration simulation was performed in the NTP ($T = 300$ K and $P = 1$ atm) ensemble. In the production MD simulation process, the SHAKE algorithm [41] was used to constrain all of the covalent bonds involving hydrogen atoms. The cut-off for calculating the short-range interactions (electrostatic and Van der Waals interactions) was set to 10 Å, and the Particle Mesh Ewald (PME) algorithm [42] was used to handle the long-range electrostatic interactions. The time step was set to 2 fs, and the snapshots were collected at an interval of 10 ps. At least 1 ns production simulation is performed in the sampling phase, but we add an evaluation of the equilibration. The root mean square deviation is evaluated during the simulation. Once the fluctuation of the root mean square deviation value of the backbone of the whole system is smaller than 1.0 Å, the simulation will be terminated. Elsewise, the simulation will proceed to the next 1 ns until the maximum duration of 10 ns. Finally, the last 1000 frames were extracted for the following binding free energy calculation.

Computational alanine mutations

We have previously developed computational mutation scanning (CMS) method [32] to predict binding affinity change in protein-organic compound system upon residue mutation. Here, we produce alanine variant structure following the method. More details can be found in the literature [43], the following is a brief protocol: (1) single-alanine substitution was performed in sequence on wild-type conformations by PyMOL software (Version 1.2, Schrödinger, LLC); (2) a standard energy minimization for relaxing mutant conformations by AMBER program and (3) a short period of MD simulation to refine the orientation of residue side chains. The unreasonable conformation of the side chains after mutation may influence the conformation of the backbone. Thus, the backbone of the protein is restrained with an elastic constant of 50 kcal·mol^{-1}·Å^{-2}, and other atoms are free to move. Energy minimization is performed by 5000 cycles of steepest descent and 5000 cycles of conjugate gradient. Subsequently, 50 ps MD simulation is performed in NPT ensemble.

Binding free energy calculation

For each protein–nucleic acid complex system, the binding free energy was calculated by the FoldX algorithm [44] based on the conformational ensemble extracted from 1 ns MD trajectory. FoldX is an empirical force field that is applicable to evaluate the binding affinity of macromolecule complex based on its three-dimensional structure. It shows good performance in computing free energies over protein–DNA structures [45] and allows calculations on structures containing RNA molecules [46]. In brief, free energy calculation by FoldX is based on the sum of empirical energy terms:

$$\Delta G = \Delta G_{\mathrm{vdw}} + \Delta G_{\mathrm{solH}} + \Delta G_{\mathrm{solP}} + \Delta G_{\mathrm{wb}} + \Delta G_{\mathrm{hbond}} + \Delta G_{\mathrm{el}} + \mathrm{T}\Delta S_{\mathrm{mc}} + \mathrm{T}\Delta S_{\mathrm{sc}} \tag{1}$$

where ΔG is the folding free energy. The terms ΔG_{vdw}, ΔG_{solH}, ΔG_{solP}, ΔG_{wb}, $\Delta G_{\mathrm{hbond}}$, ΔG_{el}, ΔS_{mc} and ΔS_{sc} represent the contributions from Van der Waals, solvation energy for apolar, polar groups, extra stabilizing free energy provided by water molecules, hydrogen bond formation, electrostatic of charged groups, the entropic cost for fixing the backbone and side chain in the folded state, respectively. For protein–nucleic acid interactions, FoldX calculates $\Delta\Delta G$ of interaction:

$$\Delta\Delta G_{\mathrm{com}} = \Delta G_{\mathrm{com}} - \left(\Delta G_{\mathrm{prot}} + \Delta G_{\mathrm{na}}\right) + \Delta G_{\mathrm{kon}} + \Delta S_{\mathrm{sc}} \tag{2}$$

$\Delta\Delta G_{\mathrm{com}}$ is the binding free energy of protein and nucleic acid molecules. The terms ΔG_{com}, ΔG_{prot}, ΔG_{na}, ΔG_{kon} and ΔS_{sc} represent the folding free energy of the complex, mono-protein, mono-nucleic acid, the effect of electrostatic interactions on the k_{on} and the loss of translational and rotational entropy for complex formation, respectively.

The binding free energy change ($\Delta\Delta G_{\varepsilon}$) as a consequence of single alanine mutation is the difference of binding affinity between the mutant ($\Delta\Delta G_{\mathrm{com,mut}}$) and wild-type ($\Delta\Delta G_{\mathrm{com,wt}}$) complexes:

$$\Delta\Delta G_{\varepsilon} = \Delta\Delta G_{\mathrm{com,mut}} - \Delta\Delta G_{\mathrm{com,wt}} \tag{3}$$

Hotspot residues are a small subset of residues that have a more significant contribution to the binding free energy than other residues. In protein–nucleic acid systems, hotspot residues can be defined operationally as those for which alanine mutations have destabilizing effects on the total binding free energy of more than 1.0 kcal·mol^{-1} [24, 31]. In most studies, researchers achieved consensus on the threshold value of 1.0 kcal·mol^{-1} to define hotspot residues [26, 47, 48]. Thus, we chose the same criterion to compare the performance of HISNAPI with the other methods. We defined hotspot residues as the ones which cause the binding free energy change of greater than 1.0 kcal·mol^{-1}.

Dynamical properties analysis

The movement stability of per residue is represented by root mean square fluctuation (RMSF). The RMSF is a measure of the deviation between the position of atom i (r_{i}) and the reference position ($r_{\mathrm{i}}^{\mathrm{ref}}$):

$$\mathrm{RMSF}_i = \left[\frac{1}{T}\sum_{\mathrm{t}_j=1}^{T}\left|r_i\ (\mathrm{t}_j) - r_i^{\mathrm{ref}}\right|^2\right]^{1/2}$$

where T is the time over the MD simulation and $r_{\mathrm{i}}^{\mathrm{ref}}$ is the reference position of atom i. Here, the reference position is the time-averaged position of atom i.

The movement correlation of pairwise residues is represented by the dynamical cross-correlation matrix (DCCM). The correlation between the atom i and atom j is defined by the following equation,

$$\mathrm{C}\,(i,j) = \left\langle \Delta r_{\mathrm{i}} \cdot \Delta r_j \right\rangle / \left\langle \Delta r_{\mathrm{i}}^{2} \right\rangle^{1/2} \left\langle \Delta r_j^{2} \right\rangle^{1/2}$$

where Δr_{i} and Δr_j are the fluctuations of atom i and j from the time-average positions, and the angle brackets represent the average over the simulation time.

Normal mode analysis (NMA) provides the information about the direction of large-amplitude movement of a protein in MD simulation. Normal mode calculation is based on the harmonic approximation of the potential energy function (V) around a minimum energy conformation:

$$\mathrm{V}\,(r_n) = \frac{1}{2} \sum_{\substack{i \\ \alpha = x,y,z}}^{N} \sum_{\substack{j \\ \beta = x,y,z}}^{N} \left.\frac{\partial^2 \mathrm{V}}{\partial_{r_{\mathrm{i}\alpha}}\partial_{r_{j\beta}}}\right|_R \upsilon_{\mathrm{i}\alpha}\upsilon_{j\beta}$$

where r is the distance between atoms i and j, R is the average distance, υ is the difference of average distance and α and β represent the direction of the motion. The routines for NMA include a Hessian construction, Newton–Raphson minimization and normal mode calculations. More details can be found in the previous literature [49]. The snapshots extracted from the simulations need to be fully minimized. Conformational optimization was conducted by a maximum of 10 000 cycles of conjugate gradient minimization with a convergence criterion of 0.0001 kcal·mol^{-1}·Å^{-1}. Then, a maximum of 200 cycles of Newton–Raphson minimization is applied, and the convergence criterion is the root-mean-square of the gradient less than 1.0 × 10^{-12} kcal·mol^{-1}·Å^{-1}. To speed up the calculation, NMA is performed on the α-carbon atoms of the protein.

Datasets and validation

To evaluate the performance of our method, we collected 23 protein–RNA and 17 protein–DNA complexes from ProNIT [33] and dbAMEPNI [34] databases, which contains 299 experimental measured binding free energy changes upon alanine mutations (Supplementary Table S1 available online at https://academic.oup.com/bib). The distribution of experimental measured binding free energy changes is shown in

Supplementary Figure S1A available online at https://academic.oup.com/bib, in which 135 residues upon alanine mutations with larger binding affinity change ($\Delta\Delta G \geq 1.0$ kcal·mol^{-1}) are considered as hotspot residues while other 164 residues are considered as neutral residues. The distribution of protein–nucleic acid complex type is shown in Supplementary Figure S1B available online at https://academic.oup.com/bib, including 5 double-stranded RNA (dsRNA), 18 single-stranded RNA (ssRNA), 12 double-stranded DNA (dsDNA) and 5 single-stranded DNA (ssDNA). HISNAPI was evaluated using the linear regression against the experimental data and the receiver operating characteristic (ROC) analysis to distinguish hotspot residues from neutral residues.

Web server configuration

We have implemented our method via a user-friendly freely available web server HISNAPI (http://chemyang.ccnu.edu.cn/ccb/server/HISNAPI/ and http://agroda.gzu.edu.cn:9999/ccb/server/HISNAPI/). HISNAPI web server runs on a Linux system computer cluster [50, 51]. The web application uses PHP, HTML and JavaScript to serve web pages [52]. The data are stored in a database implemented in MySQL [53]. Molecular structure visualization is based on NGL Viewer [54–56] web application, which runs on all modern browsers with no additional plugins.

Results

Computational workflow

The HISNAPI prediction workflow is shown in Figure 1. HISNAPI could define the interfacial residues as candidates that will undergo the following mutation scanning. The method employs MD simulation to generate the conformational ensemble of a protein–nucleic acid complex in the wild-type residue environment. A total of 1000 snapshots are collected from the equilibrated MD trajectory at regular intervals to obtain an ensemble of wild-type complex structures. Then, single-alanine mutagenesis of the wild-type ensemble is performed sequentially on mutation sites to obtain mutant conformational ensembles. Both wild-type and mutant structural conformations are refined by a standard energy minimization protocol. Following structural optimization, the average binding free energy of a conformational ensemble is calculated by the FoldX algorithm. The consequences of alanine mutations are represented by the difference in the binding free energy between the mutant and wild-type (Equation 3). According to the consensus from previous studies [30], the residues that cause large binding affinity reduction ($\Delta\Delta G_{\varepsilon} \geq 1.0$ kcal·mol^{-1}) are identified as hotspot residues. In addition, the dynamic properties of key residues could be analyzed from simulation trajectories, including the stability, correlation and direction of the residue movement.

The usage of web server

The server provides two options for the user to perform hotspot prediction, as shown in the job submission interface (Figure 2A). The ‘Site-directed mutations’ allow the user to predict the effects of alanine mutations at specific sites on the binding affinity of a protein–nucleic acid complex. The user should upload the initial structure file of the protein–nucleic acid complex or input the PDB ID collected in the Protein Data Bank database. A standard PDB-format file is recommended. The users need to input the chain ID of a single protein chain, as well as residue numbers of mutation sites in the structure file. The ‘Single-chain scanning’ allows the user to perform alanine scanning on all interfacial residues which are defined by the distance criterion. In addition to a complex structure file and a protein chain ID, users need to input a distance cut-off value. Amino acids within the distance of nucleic acids are considered mutation sites. Generally, a default parameter of 4 Å is advisable to detect all the interfacial residues. User may input any reasonable parameter value to suit her or his studies. Users are allowed to perform MD simulations with a certain ionic strength. They can either assign the concentration of sodium ion and chloride ion in the prompt boxes, or ignore the prompt boxes to neutralize the system by counter ions. Example input is present in the textbox to guide the user to input the parameters. Some extra information is optional for users, including task name, emails and passwords, which help users mark the job, receive the notice and keep the job private. A sample submission entry is available on the submission page to guide users to submit their jobs. A help page has been implemented and is accessible via the top navigation bar.

The simplified exhibition of the results page is shown in Figure 2B. The user can access the results page by the link on the Jobs web page. All the result files can be downloaded as text and image files by the link at the top of the results page. The results page contains three main panels: (1) The ‘Hotspots Identification’ panel shows the predicted hotspot residues. A figure of the protein–nucleic acid complex structure is shown, in which mutation sites are highlighted by sphere style and hotspots are colored red. The wild-type structure uploaded by the user can be visualized directly by the NGL Viewer web application. Protein and DNA/RNA molecules are shown as cartoon style, and mutation sites are highlighted in red stick style. The user can interactively operate the style of the three-dimensional structure mode to have a profile of residue packing at an atomic level. (2) The ‘Binding Free Energy Change’ panel shows the predicted change in binding free energy upon alanine mutations. Additional structural characteristics corresponding to a mutation site are analyzed, including secondary structure elements, hydrogen bonds and changes in solvent accessible surface area. The main components of the binding free energy change are also provided for analysis of dominant interactions. In addition, the energy terms are displayed in a histogram. (3) The ‘Hotspots Dynamics Analysis’ panel presents the dynamical properties focusing on the predicted hotspots, including RMSF, principal component analysis (PCA), NMA, DCCM and hydrogen bonds formation between protein and DNA/RNA molecules. RMSF represents the movement fluctuation of each residue based on the alpha-carbon atoms. The larger the RMSF value is, the more flexible the residues are. PCA describes the dominant changes in the conformational ensemble obtained by MD simulation. A PCA object stores the covariance matrix and principal modes. NMA is one of the vector quantization-based techniques used to probe large-scale motions in biomolecules. A typical application is for the prediction of functional motions in proteins. DCCM allows the identification of the correlated and anti-correlated motions of all pairwise residues. In the heatmap plot, positive values (red) represent a correlated motion between the corresponding residue pair, while negative values (blue) represent an anti-correlated motion. The value of 1 or −1 means a fully correlated and anti-correlated motion. The formation of H-bonds during the MD simulation is listed in the table. Detailed H-bonds components are recorded, including the H-bonds acceptor (Acceptor), donor and hydrogen atom in donor (Donor, DonorH), average distance (AvgDist) and angle (AvgAng).

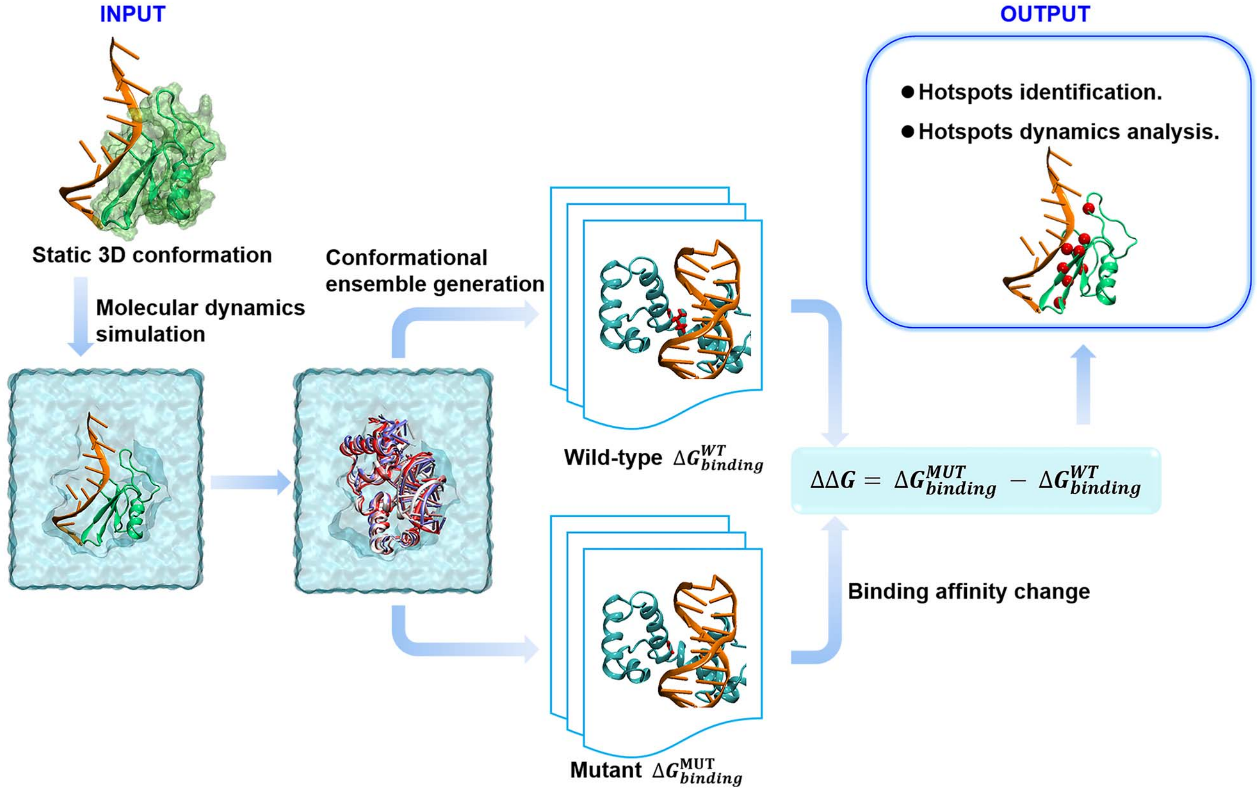

INPUT
Static 3D conformation
Molecular dynamics simulation
Conformational ensemble generation
Wild-type $\Delta G^{WT}_{binding}$
Mutant $\Delta G^{MUT}_{binding}$
$\Delta\Delta G = \Delta G^{MUT}_{binding} - \Delta G^{WT}_{binding}$
Binding affinity change
OUTPUT
● Hotspots identification.
● Hotspots dynamics analysis.

Figure 1. HISNAPI workflow. The method employs molecular simulation to generate the conformational ensemble of a protein–nucleic acid complex on the wild-type residue environment. Then, the alanine mutagenesis is performed to obtain mutant conformations. The average binding free energy change between wild-type and mutant is calculated by FoldX algorithm. In addition, dynamical behaviors of the hotspots residues are characterized from simulation trajectories.

Table 1. Statistical analysis of prediction results in protein–RNA dataset

Protein–RNA		Experimental data		Total
		Hotspot residues	Neutral residues	
Predicted data	Hotspot residues	54	9	63
	Neutral residues	13	64	77
Total		67	73	140

Table 2. Statistical analysis of prediction results in protein–DNA dataset

Protein–DNA		Experimental data		Total
		Hotspot residues	Neutral residues	
Predicted data	Hotspot residues	39	12	51
	Neutral residues	19	89	108
Total		58	101	159

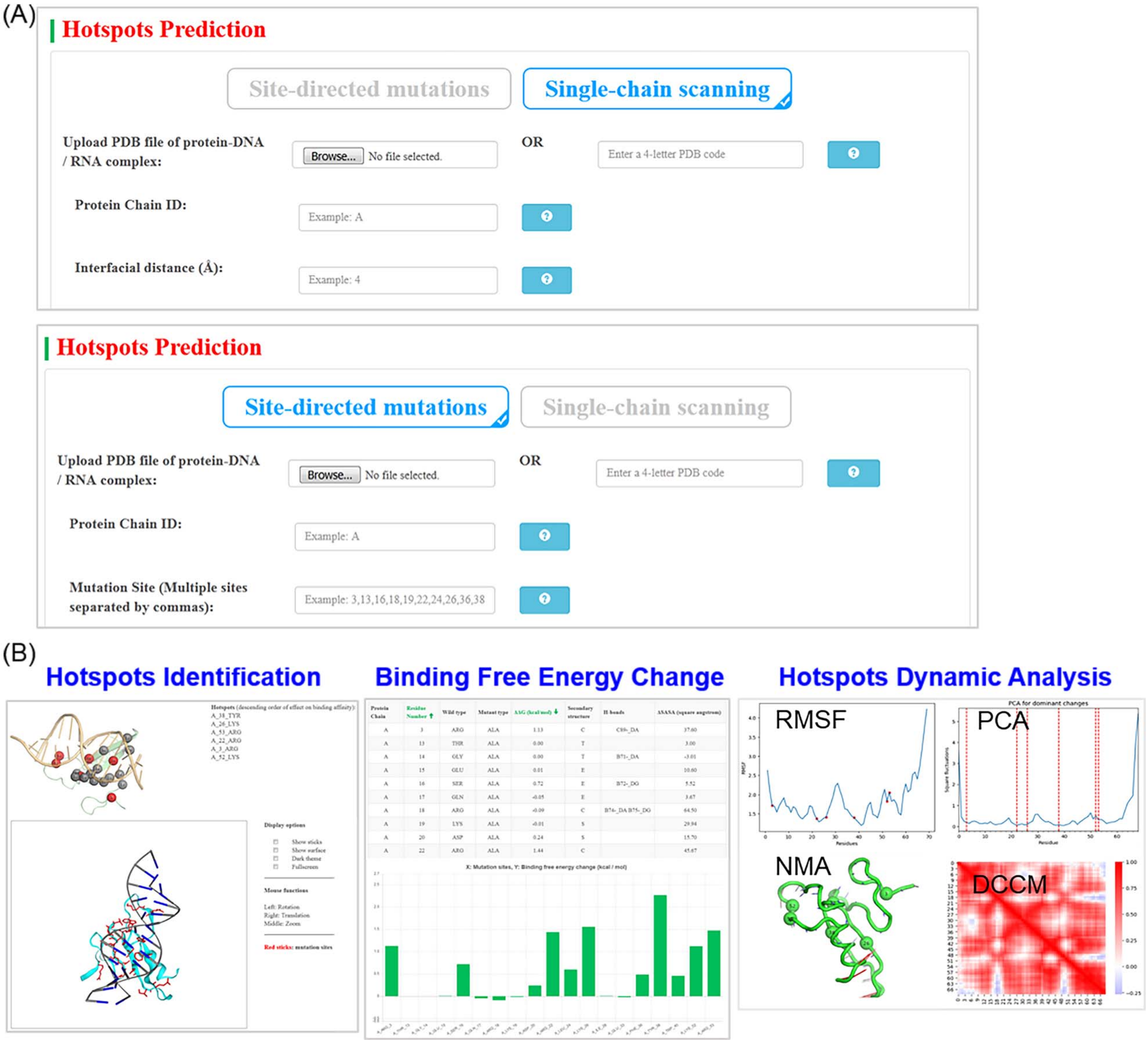
Performance

We performed ROC analysis to evaluate the performance in distinguishing hotspot residues from neutral residues. The total test dataset contains 135 hotspot residues and 164 neutral residues. HISNAPI accurately predicted 93 hotspot residues and 153 neutral residues with a true positive rate (sensitivity) of 74.4% and a true negative rate (specificity) of 87.8% (Tables 1–3). HISNAPI achieved good performance in protein–RNA and protein–DNA datasets. Figure 3A–C shows the ROC curve for classification capability. The area under the curve (AUC) value was 0.86 for the total dataset (Figure 3A), indicating the satisfactory capability of HISNAPI to distinguish different types of mutations. The AUC values for the protein–RNA and protein–DNA datasets were 0.89 and 0.81, respectively (Figure 3B and C).

Table 3. Statistical analysis of prediction results in the total dataset

Protein–nucleic acid		Experimental data		Total
		Hotspot residues	Neutral residues	
Predicted data	Hotspot residues	93	21	114
	Neutral residues	32	153	185
Total		125	174	299

A comparison against experimental data was further carried out to evaluate the performance of HISNAPI in predicting the



Hotspots Prediction

Site-directed mutations | Single-chain scanning

Upload PDB file of protein-DNA / RNA complex: Browse... No file selected. OR Enter a 4-letter PDB code

Protein Chain ID: Example: A

Interfacial distance (Å): Example: 4

Hotspots Prediction

Site-directed mutations | Single-chain scanning

Upload PDB file of protein-DNA / RNA complex: Browse... No file selected. OR Enter a 4-letter PDB code

Protein Chain ID: Example: A

Mutation Site (Multiple sites separated by commas): Example: 3,13,16,18,19,22,24,26,36,38

Hotspots (descending order of effect on binding affinity):
A_38_TYR
A_26_LYS
A_53_ARG
A_22_ARG
A_3_ARG
A_52_LYS

Display options

□ Show sticks
□ Show surface
□ Dark theme
□ Fullscreen

Mouse functions

Left: Rotation
Right: Translation
Middle: Zoom

Red sticks: mutation sites

Protein Chain	Residue Number ↑	Wild type	Mutant type	ΔΔG (kcal/mol) ↓	Secondary structure	H bonds	ΔSASA (square angstrom)
A	3	ARG	ALA	1.13	C	C89-_DA	37.60
A	13	THR	ALA	0.00	T		3.00
A	14	GLY	ALA	0.00	T	B71-_DA	-3.01
A	15	GLU	ALA	0.01	E		10.60
A	16	SER	ALA	0.72	E	B72-_DG	5.52
A	17	GLN	ALA	-0.05	E		3.67
A	18	ARG	ALA	-0.09	C	B74-_DA B75-_DG	64.50
A	19	LYS	ALA	-0.01	S		29.94
A	20	ASP	ALA	0.24	S		15.70
A	22	ARG	ALA	1.44	C		45.67

X: Mutation sites, Y: Binding free energy change (kcal / mol)

PCA for dominant changes

Figure 2. Example of job submission and results web page. (**A**) ‘Site-directed mutations’ and ‘Single-chain scanning’ are two alternative ways for hotspots prediction. (**B**) The results provided by HISNAPI consist of three aspects, including hotspots identification, binding free energy change and dynamic analysis.

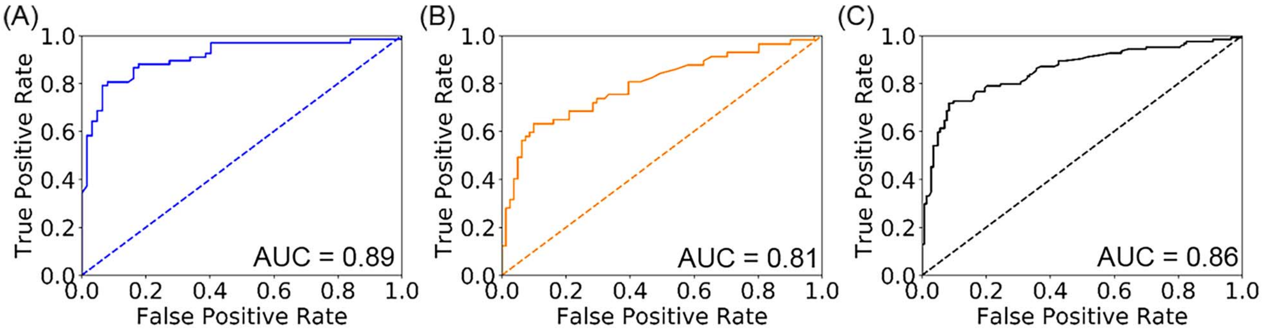


Figure 3. The ROC curve of classification of hotspots ($\Delta\Delta G \geq 1.0$ kcal·mol^{-1}) and neutral residues ($\Delta\Delta G < 1.0$ kcal·mol^{-1}) in protein–RNA (**A**), protein–DNA (**B**) and the total datasets (**C**).

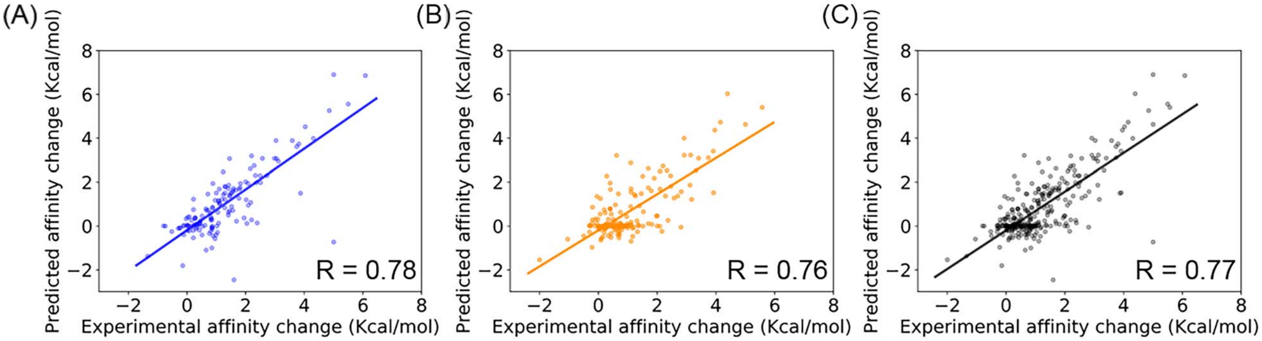


Figure 4. Regression plot between the experimental and the conformational ensemble predicted changes in binding affinity for protein–RNA (**A**), protein–DNA (**B**) and all dataset (**C**).

Table 4. Prediction performance of HISNAPI in comparison with other approaches

Methods	Principle	Datasets	Number of mutations	R[a]	AUC[b]	Dynamic analysis	Web	Computation time (min)
HISNAPI	Force field	27 protein–RNA and 13 protein–DNA complexes	299	0.77	0.86	Yes	http://chemyang.ccnu.edu.cn/ccb/server/HISNAPI/ or http://agroda.gzu.edu.cn:9999/ccb/server/HISNAPI/	~150
mCSM-NA	Machine learning	14 protein–RNA and 4 protein–DNA complexes	81	0.70	-	No	http://structure.bioc.cam.ac.uk/mcsm_na	~10
SAMPDI	Force field	13 protein–DNA complexes	105	0.72	-	No	http://compbio.clemson.edu/SAMPDI	~15
PrabHot	Machine learning	47 protein–RNA complexes	209	-	0.86	No	http://denglab.org/PrabHot/	~20
PrPDH	Machine learning	24 protein–DNA complexes	64	0.51	0.76	No	http://bioinfo.ahu.edu.cn:8080/PrPDH	~10
XGBPRH	Machine learning	15 protein–RNA complexes	58	0.66	0.81	No	https://github.com/SupermanVip/XGBPRH	~20
SPHot	Machine learning	15 protein–RNA complexes	58	0.65	0.84	No	http://bioinfo.ahu.edu.cn:8080/SPHot	~20

[a] R: Pearson correlation coefficient.
[b] AUC: The area under the curve of ROC.

binding free energy change. Figure 4 depicts the linear regression plots between the experimental and predicted data. The analysis showed a Pearson correlation coefficient of 0.78 for the protein–RNA dataset (Figure 4A) and 0.76 for the protein–DNA dataset (Figure 4B). The performance of HISNAPI seems as good in the protein–RNA dataset as in the protein–DNA dataset. For the total dataset, HISNAPI achieved a Pearson correlation coefficient of $R = 0.77$ (Figure 4C). In our previous works [43, 57], we demonstrated that compared with that based on static structures, the calculation accuracy based on conformational ensemble could be improved. Here, we also performed alanine scanning based on static structures of the protein–nucleic acid complex. The results showed a Pearson correlation coefficient of 0.41 for the total dataset, with 0.39 for the protein–RNA dataset and 0.44 for the protein–DNA dataset (Supplementary Figure S2 available online at https://academic.oup.com/bib). This outcome indicates that conformational sampling has an effect on the accuracy of the binding free energy calculation. Sufficient conformations could provide more reliable prediction results.

We finally compared the hotspot prediction performance of HISNAPI with some up-to-date and widely used approaches. These approaches use force field or machine learning-based techniques to predict the mutation effects, and are available by online web servers or stand alone. We constructed a dataset consisting of 299 mutants to test HISNAPI, while other approaches used smaller datasets (Table 4). The test datasets for HISNAPI and mCSM-NA include protein–DNA and protein–RNA complex systems. HISNAPI achieved a correlation coefficient (R) of 0.77, which demonstrates its good capability to predict the effects of alanine mutations. Except for PrPDH, which has a minimum R value of 0.51, the other approaches have R values greater than 0.6. Both HISNAPI and PrabHot have the largest AUC value of 0.86 among those methods. This only indicated the considerable ability of HISNAPI and PrabHot to distinguish hotspot residues from neutral residues. We observed that all approaches obtained satisfactory AUC values greater than 0.7. In addition, we collected a benchmark protein–RNA dataset and a benchmark protein–DNA dataset (Supplementary Table S2 available online at https://academic.oup.com/bib) to evaluate the performance of HISNAPI and the other methods. These datasets are not used to train models for the machine learning-based methods. The benchmark protein–RNA dataset consists of 15 complexes,

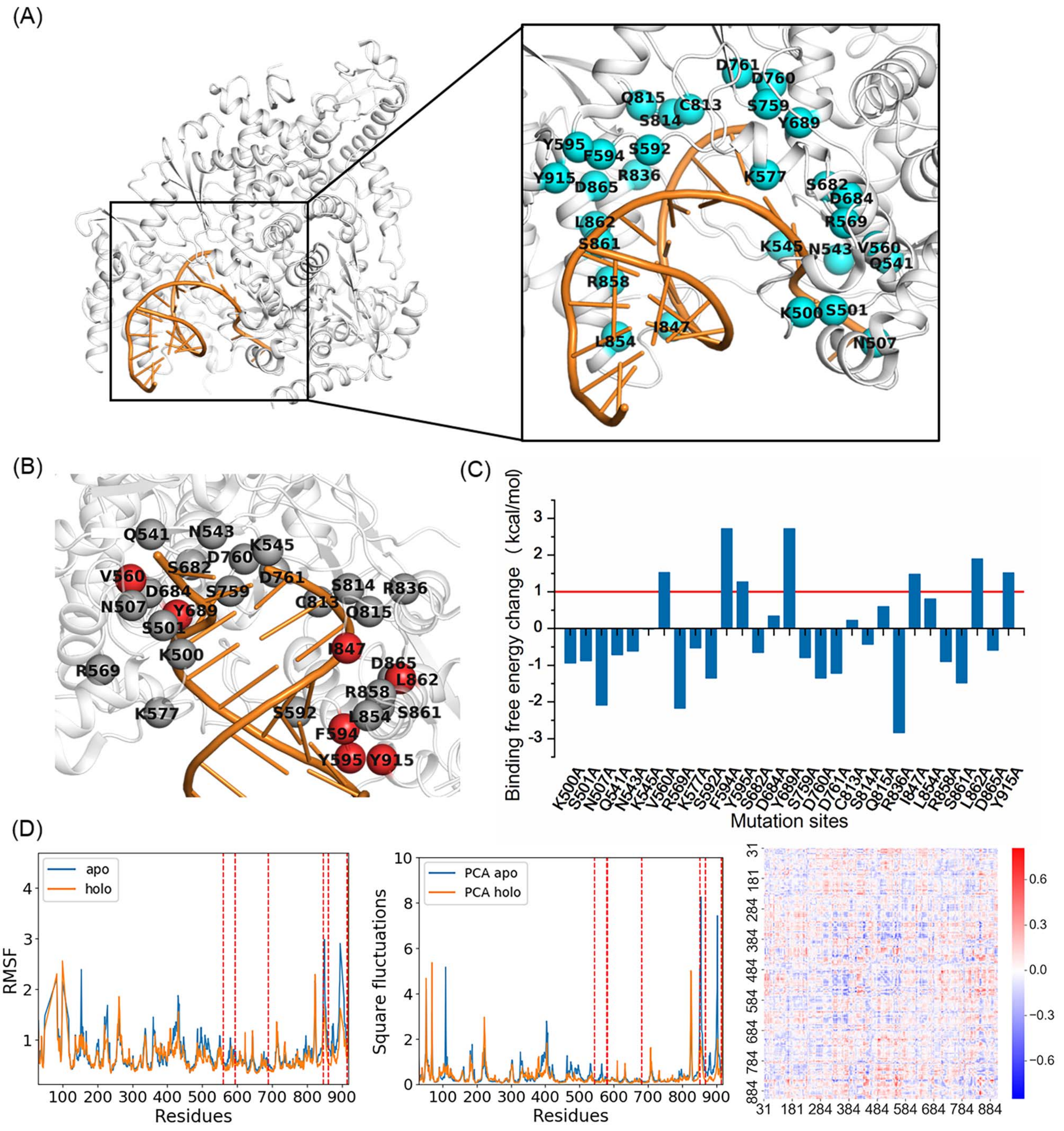


Figure 5. Case study of RdRp complex from SARS-CoV-2. (**A**) The cryo-EM structure of RdRp (white cartoon) and RNA (orange cartoon) complex (PDB ID: 7BV2). Those interfacial residues are highlighted by cyan spheres. (**B**) The hotspot residues predicted by HISNAPI. The hotspot residues are represented by red spheres, and the neutral residues by gray spheres. (**C**) The binding free energy changes upon alanine substitutions of RdRp complex from HISNAPI prediction data. (**D**) Characterization of dynamic properties of RdRp protein in apo and holo forms, including RMSF (left), PCA (middle) and DCCM (right).

a total of 99 alanine mutations. The benchmark protein–DNA dataset consists of 10 complexes, a total of 73 alanine mutations. As shown in Supplementary Table S3 and Figure S3 available online at https://academic.oup.com/bib, the prediction accuracy of HISNAPI is 0.71 for protein–RNA dataset and 0.75 for protein–DNA dataset. The AUC value is 0.84 for protein–RNA dataset and 0.79 for protein–RNA dataset. In protein–RNA system, HISNAPI and XGBPRH have the same accuracy of 0.71. The lowest accuracy achieved by SPHot is 0.58. In protein–DNA system, HISNAPI has the highest accuracy of 0.75. The other three methods achieved the accuracy above 0.6. The results prove that HISNAPI has a comparative performance in predicting protein–nucleic acid hotspot residues. HISNAPI and mCSM-NA are applicable to both protein–DNA and protein–RNA complex systems. Most of those approaches are developed exclusively for a single system. HISNAPI, mCSM-NA and SAMPDI can quantitatively predict the binding free energy changes, while PrabHot, PrPDH, XGBPRH and SPHot score the probability of whether a residue is a hotspot. XGBPRH and SPHot provide the source code, which can predict locally in batches. Among all approaches, HISNAPI is the only one for not only hotspot prediction but also dynamic behavior analysis of those hotspots. However, we still need to realize the limitations in HISNAPI. The HISNAPI method simulates the flexibility of protein–nucleic acid complexes at the cost of computational resources and time. It costs much calculation time for a large complex system. We recorded the time taken by all the methods to complete a benchmark task. The average time spent by each method to complete a job represented its calculation speed. The computational time is listed in Table 4. HISNAPI needs a longer calculation time of ∼150 minutes for a job, while other methods can complete a job in ∼30 minutes. Compared with the machine learning-based scoring methods, computational speed is one of the limitations of the structure-based computational method. Currently, HISNAPI is only capable of predicting the effects of alanine mutations and is thus not suitable for other natural amino acid mutations.

Case study

SARS-CoV-2 RNA-dependent RNA polymerase (RdRp), a vital component for the replication of the virus, is an attractive drug target for the treatment of COVID-19 disease. The cryo-EM structures of SARS-CoV-2 RdRp in apo form and in complex with a template-primer RNA and remdesivir have been reported [58]. RdRp is a so stable enzyme that no significant conformational changes between the apo and the holo form structures are observed. This work revealed the inhibition mechanism by remdesivir and provided a structural basis for designing antiviral drugs based on nucleotide analogs. We applied HISNAPI to the RdRp–RNA complex to provide a basic understanding of RNA recognition by RdRp. A total of 29 residues (except alanine and glycine) in RdRp were found to be within 4 Å of the RNA molecule (Figure 5A). HISNAPI performed alanine scanning on these interfacial residues. The results showed 7 hotspot residues (V560, F594, Y595, Y689, I847, L862 and Y915) and 22 neutral residues (Figure 5B). The binding affinity changes upon alanine substitution are shown in the histogram (Figure 5C). We observed the large reductions of binding affinity in Y595A and Y915A mutations. The results revealed that the residues Y595 and Y915 contribute mostly to the interaction between RdRp and RNA molecules.

HISNAPI also provides insights into the dynamic properties of hotspot residues in the RdRp protein. Binding site dynamics have been recognized as a vital factor that has an effect on the interaction specificity and affinity of the binding partners in drug design [59]. Figure 5D shows the routine analysis of residue movement, including RMSF (left), PCA (middle) and DCCM (right). In the RMSF plot, the RMSF value reports on the fluctuations that take place on a per residue basis during the simulation. The movements of RdRp protein in apo form (blue) and in holo form (orange) are recorded. The red dotted lines highlight the positions of the predicted hotspot residues. The larger the RMSF value is, the more flexible the residue is. RMSF analysis revealed that the whole structures of RdRp in apo and holo forms share a similar tendency for residue fluctuation. The regions around the RNA binding interface decreased the fluctuation upon RNA recognition. We observed that the hotspot residues tend to be located in locally stable regions with smaller RMSF values than that of the flank. The movements of the hotspots on V560, F594, Y595 and Y689 were not affected by RNA binding. The PCA plot shows the dominant changes in the conformation ensemble during the simulation. The conformation space sampled by MD simulation is used to build and diagonalize the covariance matrix to determine the principal modes of structural variations. The red dotted lines highlight the positions of the predicted hotspot residues. The apo form (blue) RdRp has dramatic conformational changes in some regions compared with the holo form (orange). However, these hotspot residues of V560, F594, Y595 and Y689 in both forms maintain stable conformations. This revealed the conformational stability on these sites, which is consistent with the RMSF analysis. 'Stability patches' have been reported to exist on protein surfaces, regulating protein recognition interactions [60]. The stable backbone of hotspot residues may contribute to their role in facilitating the binding of small ligands to the binding sites. DCCM describes the movement correlation of all pairwise residues. The DCCM plot shows the difference in the correlation coefficient between the holo form and the apo form RdRp. Positive values (red) represent the motion that becomes more correlated, while negative values (blue) represent the motion that becomes more anti-correlated. In DCCM plot, we observed that the changes of movement correlation are low-scale and occurred in small local regions throughout the RdRp protein structure. This indicates that RNA binding had no substantial effect on the correlation of protein motion. It may be explained by the structural stability of RdRp in response to the perturbation of RNA binding. Protein flexibility plays a vital role in molecular recognition. The dynamic behaviors may help to predict the binding interaction at the residue level. Computational approaches to identify hotspot residues and describe their dynamics provide a means to reveal potential binding pockets to expand the possibility for improving drug design.

Conclusions

In our previous work, we developed the PIIMS (Protein Interface *In silico* Mutation Scanning) web server for predicting the hotspot residues in protein–protein complex interfaces and the effects of hotspot mutations on protein–protein interactions. The calculation strategies of HISNAPI and PIIMS are both based on the CMS protocol. However, HISNAPI is designed to characterize the dynamic properties of hotspot residues, while PIIMS is designed to predict the effects of hotspot mutations. HISNAPI utilizes an empirical force field FoldX to evaluate the binding free energy, and PIIMS uses the MM/PBSA method.

HISNAPI was developed to predict the hotspot residues in protein–nucleic acid interfaces and provide dynamic information of the hotspots. Considering the flexibility of interfacial residues, HISNAPI evaluates the binding affinity based on conformational ensemble by conformational sampling rather than static structure. And HISNAPI utilizes an empirical force field FoldX to evaluate the binding affinity, which is applicable not only to the protein–DNA system but also to the protein–RNA system. The method achieved a correlation of up to 0.77 between experimental and predicted data and was able to classify the large and small effects upon alanine mutations with an AUC value of 0.86, showing its high capability to identify hotspot amino acids. HISNAPI provides a way to connect structure, dynamics and function in macromolecule binding processes by extracting dynamic properties from molecular simulations. MD simulations allow the motion of a protein to be predicted over time. It is a powerful approach to study the dynamic properties at the atomic level. Simulation time is a main limitation for accurate analysis of protein motion. A basic assumption is made in our method that affects the applicability of HISNAPI and the interpretation of the results. It is assumed that there are no dramatic conformational changes in a protein–nucleic acid complex. Nanosecond time-scale MD simulations are possible to describe the dominant conformational space for a rigid protein system. For a flexible or disordered protein system, HISNAPI could not well describe the dynamic properties. We believe that HISNAPI is a multifunctional tool for mutagenesis research to guide experimentation, shedding light on the mechanism of protein–nucleic acid binding at the molecular level.

Key Points

- Accurate and fast predictions of hotspots on protein–nucleic acid interfaces are essential for understanding the mechanisms of protein–nucleic acid interactions.
- We have made our method available as a webserver HISNAPI for prediction of the hotspot residues on protein–nucleic acid interface.
- Besides the superior performance in prediction of hotspot residues, HISNAPI has unique capability to describe the dynamic properties of hotspots, leading to the connection between structure, dynamics and function.
- HISNAPI is free access for public without registration.

Supplementary Data

Supplementary data are available online at *Briefings in Bioinformatics*.

Funding

National Key R&D Program (2018YFD0200100); National Natural Science Foundation of China (21772059, 91853127 and 31960548); Program of Introducing Talents of Discipline to Universities of China (111 Program, D20023); Frontiers Science Center for Asymmetric Synthesis and Medicinal Molecules, Department of Education, Guizhou Province (Qianjiaohe KY number [2020]004).

References

1. Kuznetsova SA, Oretskaya TS. Structure and function analysis of protein–nucleic acid complexes. *Russ Chem Rev* 2016;**85**:445–63.
2. Puglisi JD, Doudna JA. Nucleic acids and their protein partners. *Curr Opin Struct Biol* 2008;**18**:279–81.
3. Moreira IS, Fernandes PA, Ramos MJ. Hot spots–a review of the protein-protein interface determinant amino-acid residues. *Proteins* 2007;**68**:803–12.
4. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* 2002;**99**:14116–21.
5. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* 2002;**12**: 14–20.
6. Keskin O, Ma B, Nussinov R. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* 2005; **345**:1281–94.
7. Ji B, Liu S, He X, *et al.* Prediction of the binding affinities and selectivity for CB1 and CB2 ligands using homology Modeling, molecular docking, molecular dynamics simulations, and MM-PBSA binding free energy calculations. *ACS Chem Neurosci* 2020;**11**:1139–58.
8. Li M, Cong Y, Li Y, *et al.* Insight into the binding mechanism of p53/pDIQ-MDMX/MDM2 with the interaction entropy method. *Front Chem* 2019;**7**:33.
9. Duan J, Hu C, Guo J, *et al.* A molecular dynamics study of the complete binding process of meropenem to New Delhi metallo-beta-lactamase 1. *Phys Chem Chem Phys* 2018;**20**: 6409–20.
10. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci U S A* 2005;**102**:6825–30.
11. Lee TS, Hu Y, Sherborne B, *et al.* Toward fast and accurate binding affinity prediction with pmemdGTI: an efficient implementation of GPU-accelerated thermodynamic integration. *J Chem Theory Comput* 2017;**13**:3077–84.
12. Wang E, Sun H, Wang J, *et al.* End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev* 2019;**119**:9478–508.
13. Lambert M, Jambon S, Depauw S, *et al.* Targeting transcription factors for cancer treatment. *Molecules* 2018;**23**:1479.
14. Fontaine F, Overman J, Moustaqil M, *et al.* Small-molecule inhibitors of the SOX18 transcription factor. *Cell Chem Biol* 2017;**24**:346–59.
15. Alonso N, Guillen R, Chambers JW, *et al.* A rapid and sensitive high-throughput screening method to identify compounds targeting protein-nucleic acids interactions. *Nucleic Acids Res* 2015;**43**:e52.
16. Seo MH, Park J, Kim E, *et al.* Protein conformational dynamics dictate the binding affinity for a ligand. *Nat Commun* 2014;**5**:3724.
17. Zerbe BS, Hall DR, Vajda S, *et al.* Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J Chem Inf Model* 2012;**52**:2236–44.
18. Yang J-F, Wang F, Chen Y-Z, *et al.* LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. *Brief Bioinform* 2019;**21**:2206–18.
19. Pargellis C, Tong L, Churchill L, *et al.* Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* 2002;**9**:268–72.
20. Tuszynska I, Bujnicki JM. DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. *BMC Bioinformatics* 2011;**12**:348.
21. Huang SY, Zou X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res* 2014; **42**:e55.
22. Li H, Huang Y, Xiao Y. A pair-conformation-dependent scoring function for evaluating 3D RNA-protein complex structures. *PLoS One* 2017;**12**:e0174662.
23. Yang W, Deng L. PreDBA: a heterogeneous ensemble approach for predicting protein-DNA binding affinity. *Sci Rep* 2020;**10**:1278.
24. Zhang SJ, Zhao L, Xia JF. SPHot: prediction of hot spots in protein-RNA complexes by protein sequence information and ensemble classifier. *IEEE Access* 2019;**7**:104941–6.
25. Zhang Q, Shen Z, Huang D-S. Modeling in-vivo protein-DNA binding by combining multiple-instance learning with a hybrid deep neural network. *Sci Rep* 2019;**9**:8484.
26. Deng L, Sui Y, Zhang J. XGBPRH: prediction of binding hot spots at protein(−)RNA interfaces utilizing extreme gradient boosting. *Genes (Basel)* 2019;**10**:242.
27. Munteanu CR, Pimenta AC, Fernandez-Lozano C, *et al.* Solvent accessible surface area-based hot-spot detection methods for protein-protein and protein-nucleic acid interfaces. *J Chem Inf Model* 2015;**55**:1077–86.

28. Pires DEV, Ascher DB. mCSM-NA: predicting the effects of mutations on protein-nucleic acids interactions. *Nucleic Acids Res* 2017;**45**:W241–6.
29. Peng Y, Sun L, Jia Z, *et al*. Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver. *Bioinformatics* 2018;**34**:779–86.
30. Pan Y, Wang Z, Zhan W, *et al*. Computational identification of binding energy hot spots in protein-RNA complexes using an ensemble approach. *Bioinformatics* 2018;**34**: 1473–80.
31. Zhang S, Zhao L, Zheng CH, *et al*. A feature-based approach to predict hot spots in protein-DNA binding interfaces. *Brief Bioinform* 2020;**21**:1038–46.
32. Hao GF, Yang GF, Zhan CG. Computational mutation scanning and drug resistance mechanisms of HIV-1 protease inhibitors. *J Phys Chem B* 2010;**114**:9663–76.
33. Kumar MD, Bava KA, Gromiha MM, *et al*. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res* 2006;**34**: D204–6.
34. Liu L, Xiong Y, Gao H, *et al*. dbAMEPNI: a database of alanine mutagenic effects for protein-nucleic acid interactions. *Database (Oxford)* 2018;**2018**:bay034.
35. Ofran Y, Rost B. Analysing six types of protein-protein interfaces. *J Mol Biol* 2003;**325**:377–87.
36. Case DA, Cheatham TE, 3rd, Darden T, *et al*. The amber biomolecular simulation programs. *J Comput Chem* 2005;**26**:1668–88.
37. Maier JA, Martinez C, Kasavajhala K, *et al*. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;**11**: 3696–713.
38. Aytenfisu AH, Spasic A, Grossfield A, *et al*. Revised RNA dihedral parameters for the amber force field improve RNA molecular dynamics. *J Chem Theory Comput* 2017;**13**: 900–15.
39. Galindo-Murillo R, Robertson JC, Zgarbova M, *et al*. Assessing the current state of amber force field modifications for DNA. *J Chem Theory Comput* 2016;**12**:4114–27.
40. Price DJ, III CLB. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* 2004;**121**: 10096–103.
41. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;**23**:327–41.
42. Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;**98**:10089–92.
43. Wu FX, Wang F, Yang JF, *et al*. AIMMS suite: a web server dedicated for prediction of drug resistance on protein mutation. *Brief Bioinform* 2018;**21**:318–28.
44. Schymkowitz J, Borg J, Stricher F, *et al*. The FoldX web server: an online force field. *Nucleic Acids Res* 2005;**33**:W382–8.
45. Blanco JD, Radusky L, Climente-Gonzalez H, *et al*. FoldX accurate structural protein-DNA binding prediction using PADA1 (protein assisted DNA assembly 1). *Nucleic Acids Res* 2018;**46**:3852–63.
46. Delgado J, Radusky LG, Cianferoni D, *et al*. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* 2019;**35**:4168–9.
47. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol* 1998;**280**:1–9.
48. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science* 1995;**267**:383–6.
49. Wang J, Hou T. Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *J Chem Inf Model* 2012;**52**:1199–212.
50. Wang F, Wu F-X, Li C-Z, *et al*. ACID: a free tool for drug repurposing using consensus inverse docking strategy. *J Chemother* 2019;**11**:73.
51. Hao G-F, Jiang W, Ye Y-N, *et al*. ACFIS: a web server for fragment-based drug discovery. *Nucleic Acids Res* 2016; **44**:W550–6.
52. Wang M-y, Wang F, Hao GF, *et al*. FungiPAD: a free web tool for compound property evaluation and fungicide-likeness analysis. *J Agric Food Chem* 2019;**67**:1823–30.
53. Wang F, Yang J-F, Wang M-Y, *et al*. Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. *Sci Bull* 2020;**65**:1184–91.
54. Contessoto VG, Cheng RR, Hajitaheri A, *et al*. The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome. *Nucleic Acids Res* 2020;**48**:gkaa818.
55. Rose AS, Bradley AR, Valasatava Y, *et al*. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* 2018;**34**:3755–8.
56. Rose AS, Hildebrand PW. NGL viewer: a web application for molecular visualization. *Nucleic Acids Res* 2015;**43**:W576–9.
57. Wu F, Zhuo L, Wang F, *et al*. Auto in Silico ligand directing evolution to facilitate the rapid and efficient discovery of drug lead. *iScience* 2020;**23**:101179.
58. Yin W, Mao C, Luan X, *et al*. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 2020;**368**:1499–504.
59. Stank A, Kokh DB, Fuller JC, *et al*. Protein binding pocket dynamics. *Acc Chem Res* 2016;**49**:809–15.
60. Kuttner YY, Engel S. Protein hot spots: the islands of stability. *J Mol Biol* 2012;**415**:419–28.