

# KFC Server Manual

## ***Introduction***

The KFC Server is a web-based implementation of the KFC model—a machine learning approach for predicting binding hot spots, or the subset of residues that account for most of a protein interface’s binding free energy. The server facilitates the automated analysis of a given protein interface and the visualization of its hot spot predictions. For each residue within the interface, the KFC Server characterizes the local structural environment surrounding the residue, compares it to known environments of experimentally determined hot spots, and predicts if the residue is a hot spot. The user can visualize the results using an interactive job viewer after the computational analysis is complete. In addition to standard molecular viewer functionality, the job viewer allows the user to quickly highlight predicted hot spots and surrounding structural features using a graphical interface.

## ***The KFC Model***

The KFC model is comprised of two decision tree-based models: *K-FADE* (based on shape specificity features), and *K-CON* (based on biochemical contacts such as intermolecular hydrogen bonds and atomic contacts). Using a data set of experimental alanine-scanning mutations, each model is trained to recognize local structural environments that are indicative of hot spots. For this work, hot spots are defined as mutations associated with a change in binding energy ( $\Delta\Delta G$ ) greater than 2 kcal/mol. The following journal article provides a complete discussion of the development and performance of the KFC Model:

Darnell, S.J., Page, D., Mitchell, J.C. (2007) “An automated decision-tree approach to predicting protein interaction hot spots.” *Proteins*, 68(4):813-23.

## ***Running a KFC Analysis***

### **Registration and Login**

Users must register prior to submitting jobs to any of the tools hosted by the Mitchell Lab. Personal information is only used to contact users when their analysis is complete; it will not be shared. To register, enter a unique user name and email address on the registration page, then click the submit button. An error message will display if the selected user name is in use by another user.

Once registered, a user may login to the server. Although login is not required to submit jobs, it allows a user to view their personal jobs in the job viewer. By default, a login will expire after two weeks; however, a user may manually logoff as well.










### **Submitting a Job**

Before the KFC analysis can begin, a user must provide the structure of a protein complex and define the interface to analyze. A protein interface is the region between the two binding partners, where each partner is comprised of one or more proteins. Files that do not contain a bound complex are unlikely to yield useful results. In addition, model structures containing many clashes may vastly overestimate the number of hot spots. Finally, KFC is able to analyze structures containing proteins and DNA/RNA but not other types of molecules. Please remove these from your PDB file before submitting to the server.

# Mitchell Lab - KFC Server

Home **Submit Job** Job Queue Job Viewer | Login Register

Note: All fields are Required

Username:  <input type="text" value="username"/>	<b>AND</b>	Email Address:  <input type="text" value="email@myschool.edu"/>	<b>Help</b> Please <b>register</b> before proceeding. Registration is free and allows notification when your job is complete.  Place your mouse over the  icon for information about each field.
Job Name:  <input type="text" value="Test"/>		<input type="checkbox"/> Private Job 	
Upload Complex:  <input type="button" value="Choose File"/> no file selected	<b>OR</b>	PDB Code:  <input type="text" value="1XXX"/>	
Protein 1 Chainlist:  <input type="text" value="ABCD"/>		Protein 2 Chainlist:  <input type="text" value="EFGH"/>	

To analyze an interface, enter the following information on the submission page and click the submit button:

- A registered user name and email address (e.g. “username” and [email@myschool.edu](mailto:email@myschool.edu))
- A short name for the job (e.g. “Test”)
- A protein complex structure (see below) or PDB Code (e.g. 1XXX)
- The chain identifiers defining binding partner #1 (e.g. ABCD)
- The chain identifiers defining binding partner #2 (e.g. DEFG)

Typically, users will want to analyze structures found in the Protein Data Bank (PDB); however, the KFC Server will accept any PDB formatted coordinate file, such as structures generated from protein docking or other molecular modeling techniques. To use a structure from the PDB, enter the four character PDB code in the appropriate field. To upload a structure from your computer, click the browse button and select the file on your computer.

## Job Queue

Upon submission, the task will enter the job queue and wait for processing. The queue displays the current status for each submitted job (Queued, Active, View Results, or Error), and provides links to KFC input and output files. After processing begins, a typical KFC analysis finishes within two minutes. When the task is complete, an email is sent to the user with their KFC hot spot predictions or an error message. If the job finishes successfully, the status field will contain a link to the interactive job viewer. Jobs that end in error are described by the following error codes:

### Error 2:

#### **Error occurred while calculating shape specificity**

*This most likely occurs because the chain ID's you have provided are incorrect and do not lead to a valid interface. Check the error file to see if it says “No atoms found.”*

### Error 3:

#### **Error occurred while calculating biochemical contact**

*This most likely occurs because there are atoms or groups that KFC does not recognize. Please make sure your PDB file contains only standard amino acids or nucleotides with atoms C, N, O, S, H. The error can also occur if the atom type is shifted into the wrong column. Check the error file to see if it complains of an unrecognized item and delete this.*

**Error 4:** **Error occurred while predicting hot spots**  
*This is unlikely to occur.*

**Error 5:** **Error occurred while deleting temporary files**  
*This is unlikely to occur.*

Users can access KFC input, output, and error files by clicking on a job's identification number. Most errors are caused by non-standard amino acids or ligands incorrectly labeled as ATOM records within the PDB coordinate file. If possible, the user should resolve the inconsistencies in the file and submit a new job. If subsequent jobs still end in error, users can contact [admin@mitchell-lab.org](mailto:admin@mitchell-lab.org) for assistance.

## Format of Hot Spot Predictions

Chain	Residue	ResID	K-FADE Class	K-FADE Conf	K-CON Class	K-CON Conf
A	LYS	27	Hotspot	0.72	Hotspot	0.60
A	TRP	35	-----	1.00	-----	1.00
A	PHE	56	Hotspot	0.66	-----	0.73
D	TYR	29	Hotspot	0.81	Hotspot	0.71
D	ASN	33	-----	0.64	Hotspot	0.50
D	LEU	34	-----	0.81	-----	0.90

*Chain:* Chain identifier from PDB file  
*Residue:* Amino acid name  
*ResID:* Residue number from PDB file  
*K-FADE (or K-CON) Class:* Predicted classification  
*K-FADE (or K-CON) Conf:* Confidence of prediction (see below)

The score for K-FADE and K-CON indicates the confidence of the prediction, where its worst value is 0 and its best value is 1.

## Using the KFC Viewer

The job viewer has two major components: a molecular viewer on the left, and a control panel on the right. Users can directly interact with the molecular viewer or use the control panel to affect the display. Each component is described in more detail below.

### Control Panel: FADE Shape Markers

KFC uses the Fast Atomic Density Evaluator (FADE) to analyze the shape specificity within a protein-protein interface. Users can highlight different degrees of shape specificity clicking on the different color-coded checkboxes.

*Well-matched:* Red, Orange  
*Flat surfaces:* Yellow, Green  
*Mismatched:* Violet, Blue

# KFC - 1brs

Home Submit Job Job Queue Job Viewer | Login Register

The screenshot displays the KFC web interface. On the left is a 3D ribbon representation of a protein complex, with one chain highlighted in yellow and others in light blue. Below the structure are buttons for 'Open Console', 'PDB File', 'Jmol Help', and 'KFC Help'. On the right, there are three main control panels: 1. 'FADE Shape Markers' with a 'Mismatch' to 'Match' gradient bar and a row of colored squares. 2. 'Display Controls' with 'Save' and 'View' buttons (1-4), and dropdown menus for 'Background: Gray', 'Style: Cartoons', 'Color: Chain', and 'Surface: None', plus a 'Show Selection' checkbox. 3. 'Interface and Hot Spots' with two rows of chain controls. The first row includes 'A' (checked) and residues LYS27:A, TRP35:A, ALA37:A, SER38:A (highlighted), and PHE56. The second row includes 'D' (checked) and residues TYR29:D, TYR30:D, GLY31:D, ASN33:D (highlighted), and LEU34. Each chain control has a checkbox and a subset selection menu.

## Control Panel: Display Controls

These controls alter the appearance of the selected atoms. By default, KFC selects all protein atoms in the complex. Advanced users may change the atom selection by using the Jmol scripting language. Selecting the “Show Selection” checkbox will highlight the current atom selection.

<i>Background:</i>	Change the color of the background
<i>Style:</i>	Change the representation of selected backbone or side-chain atoms
<i>Color:</i>	Change the color of the selected backbone or side-chain atoms
<i>Surface:</i>	Add a molecular surface to the selected atoms

Additionally, users can save up to four different views of their session. Clicking one of the “Save” buttons will record the current state of the display. Clicking the appropriate “View” button will restore the viewer to the saved state.

## Control Panel: Interface and Hot Spots

Each distinct chain produces a unique group in the interface display. The chain’s name is displayed in the leftmost box. The checkbox beneath the chain name toggles whether the chain is displayed or hidden. The popup menu determines which subset of atoms is selected for action by the Display Controls (See Caveat #1).

This panel controls the display of each interface residue and predicted hot spot, and provides summary information about each residue. The residues are separated by protein chain, and their appearances are controlled by a set of three checkboxes (one set for each residue).

*Checkbox #1:* Highlight the residue with space filling  
*Checkbox #2:* Show the residue using sticks  
*Checkbox #3:* Add a surface around the residue (see Caveat #2)

The coloring within each table cell also encodes information about the residue.

*Background color:* Chemical type of amino acid  
(gray = hydrophobic, yellow = polar, red = acidic, blue = basic)

*Highlight around name:* Classification by KFC  
(pink = predicted hot spot, white = interface residue)  
*Note: hold the mouse over a name to see KFC scores for that hot spot*

## Control Panel: Miscellaneous Buttons

*Open Console:* Opens the console, uses Rasmol commands (See Caveat #1)  
*PDB File:* Opens the PDB file used by the molecular viewer  
*Jmol Help:* Opens the documentation for Jmol  
*KFC Help:* Opens the KFC Server instruction manual (this document)

## Molecular Viewer: Jmol

Jmol is the molecular viewer used throughout the Mitchell Lab website. It is an applet written in Java, so users must enable Java and Javascript in their web browsers in order to use the KFC Server. Also, Windows users may need to install the most current Sun Java Runtime Environment (JRE) in order to use Jmol.

Jmol is extensively documented, so we direct users to the Jmol website for information about its use.

*Jmol Documentation:* <http://jmol.sourceforge.net/docs>  
*Jmol Wiki:* <http://wiki.jmol.org> (includes tutorials, such as "Using the Mouse")  
*Scripting Documentation:* <http://chemapps.stolaf.edu/jmol/docs> (for use with the console)

## Caveats

1. If you use the console to make selections and change displays, the selections shown in the Control Panel may no longer be accurate. Actions taken using the console override any mouse-driven selection and display controls.
2. When one translucent sidechain surface is deactivated, they all deactivate. When a new one is activated, all selected surfaces reappear. This seems to be a limitation with the Jmol scripting language (or at least our skill in manipulating it.)

## Publications

Please cite the following in any work that uses the KFC Server: S. J. Darnell, D. Page, and J. C. Mitchell. Automated Decision-Tree Approach to Predicting Protein-Protein Interaction Hot Spots. *Proteins*, 68(4): 813-823, 2007.