Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

Insights into the allosteric regulation of Syk association with receptor ITAM, a multi-state equilibrium

Chao Feng and Carol Beth Post*

Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology, Purdue Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, USA. E-mail: cbp@purdue.edu

1. The distinct behavior of residues from different SH2 domains

The basis of using resonances from the same domain as independent observations of the domain binding is that the residues within the same SH2 domain showed similar binding curves during titration while resonances from different SH2 domains displayed different ligand-concentration dependence.

Fig. S1a shows titration curves for resonances from the two SH2 domains of tSH2 for the case of (N)SH2 and addition of C-IHP, or (C)SH2 and the addition of N-IHP, corresponding to the contacts present in the tSH2 association with dp-ITAM (see Fig. 1 in main text). The (C)SH2-domain residues show similar titration curves, but differ from (N)SH2-domain residues, and with higher affinity than the (N)SH2-domain residues as indicated by the steeper slope of the curve. Furthermore, the titration curve for the same residue was similar for all three tSH2 constructs (Fig. S1b), suggesting the individual SH2 domain affinity is not changed significantly by linker phosphorylation (linker status).

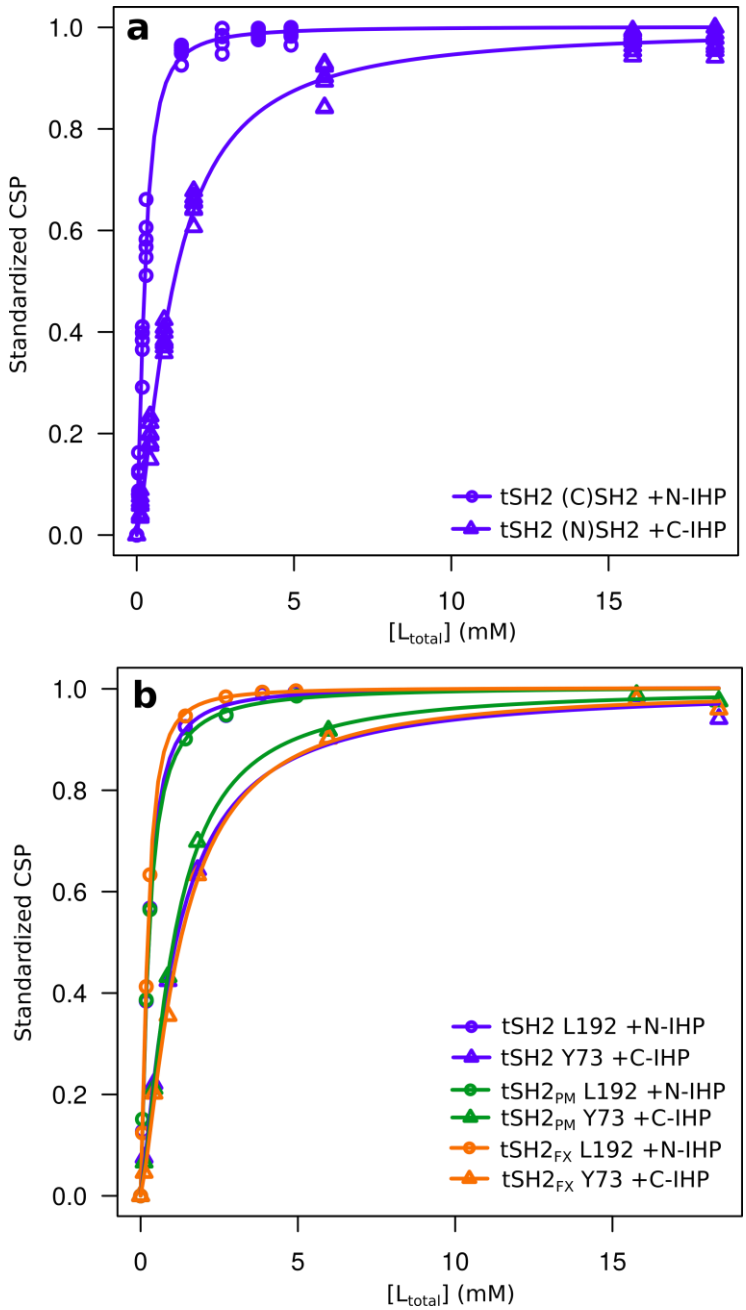

Fig. S1 (**a**) standardized CSP (chemical shift perturbation) titration plot for resonances from residues in each domain of tSH2 (0.3 mM) upon addition of the IHP peptides. These residues correspond to the contacts of tSH2 with dp-ITAM (PDB ID 1A81). Residues within the same domain showed similar behavior while residues from different domains behaved differently. Data points: values of individual resonances from each domain. Lines: concentration-dependence curves fitted with the averaged values of residues from the same domain. (**b**) Standardized CSP plot for an (N)SH2 residue (Y73) and a (C)SH2 residue (L192) binding with the matched IHP peptide in three tSH2 constructs (0.3 mM). The same residue behaved similarly in all three tSH2 constructs. Data points: values of the indicated resonance for tSH2, $tSH2_{PM}$ or $tSH2_{FX}$. Lines: concentration-dependence curves fitted for the indicated residues in each construct.

2. The 4-state modeling for tSH2 / IHP complexes

2a. Mathematical description

A tSH2 protein molecule contains two binding domains: (N)SH2 and (C)SH2. The binding interaction between tSH2 and an IHP (ITAM half peptide, which contains only 1 pY residue) is modeled with the scheme in Fig. 4 (see main text).

Based on this scheme, the equilibrium between the (N)SH2 domain unbound and bound states is characterized by the equilibrium dissociation constant, K_N,

$$K_N = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{N^LC}]} = \frac{[\mathrm{NC^L}][\mathrm{L}]}{[\mathrm{N^LC^L}]} \tag{1}$$

The equilibrium between the (C)SH2 domain unbound and bound states is characterized by the equilibrium dissociation constant, K_C,

$$K_C = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{NC^L}]} = \frac{[\mathrm{N^LC}][\mathrm{L}]}{[\mathrm{N^LC^L}]} \tag{2}$$

The total concentration of protein and ligand of each titration point is known and can be expressed as

$$[\mathrm{P_{total}}] = [\mathrm{NC}] + [\mathrm{N^LC}] + [\mathrm{NC^L}] + [\mathrm{N^LC^L}] \tag{3}$$

$$[\mathrm{L_{total}}] = [\mathrm{L}] + [\mathrm{N^LC}] + [\mathrm{NC^L}] + 2[\mathrm{N^LC^L}] \tag{4}$$

From eqn (1-4), the relationship between [L], $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_N and K_C is

$$\begin{aligned}&[\mathrm{L}]^3 + [\mathrm{L}]^2(2[\mathrm{P_{total}}] - [\mathrm{L_{total}}] + K_N + K_C)\\ &+ [\mathrm{L}]\big((K_N + K_C)[\mathrm{P_{total}}] - (K_N + K_C)[\mathrm{L_{total}}] + K_N K_C\big) - [\mathrm{L_{total}}]K_N K_C = 0\end{aligned} \tag{5}$$

With given values of $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_N and K_C, the value of [L] can be solved analytically with Maxima (a Computer Algebra System, version 5.32.1, http://maxima.sourceforge.net). Once [L] is known, the concentrations of other species in the system could be determined as follows based on eqn (1-4):

$$[\mathrm{NC}] = \frac{K_C K_N [\mathrm{P_{total}}]}{[\mathrm{L}]^2 + (K_C + K_N)[\mathrm{L}] + K_C K_N} \tag{6}$$

$$[\mathrm{N^LC}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_N} \tag{7}$$

$$[\mathrm{NC^L}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_C} \tag{8}$$

$$[\mathrm{N^LC^L}] = \frac{[\mathrm{L}]^2[\mathrm{NC}]}{K_N K_C} \tag{9}$$

2b. Titration-curve analysis

The NMR 2D ^{1}H-^{15}N HSQC peaks of IHP complexes generally showed fast or fast-intermediate exchange behavior during titration experiments, and thus the position of peak center between the resonance frequency for the unbound protein and fully saturated protein reflects the value of fraction bound of each domain. The bound fraction of (N)SH2, f_N, and (C)SH2, f_N, are defined as

$$f_N = \frac{[\mathrm{N^L C}] + [\mathrm{N^L C^L}]}{[\mathrm{P_{total}}]} \tag{10}$$

$$f_C = \frac{[\mathrm{N C^L}] + [\mathrm{N^L C^L}]}{[\mathrm{P_{total}}]} \tag{11}$$

Values of f_N and f_C were obtained from eqn (5-11) with given values of $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_N and K_C. Noteworthy, eqn (10-11) can be transformed to the following two equations for the 4-state model:

$$f_N = \frac{1}{1 + \frac{K_N}{[\mathrm{L}]}} \tag{12}$$

$$f_C = \frac{1}{1 + \frac{K_C}{[\mathrm{L}]}} \tag{13}$$

Eqn (12-13) assume no coupling and the two domains bind independently to an IHP ligand, and a hyperbolic binding curve is obtained when plotting the value of fraction bound protein against [L], where [L] is calculated with the fitted equilibrium dissociation constants.

The titration-curve fitting procedure is described by the flow chart in Fig. S2. In this procedure:

1) The chemical shift perturbation (CSP) value of the amide resonance of each residue at each titration point was experimentally determined and denoted as *d* value [$d = \sqrt{(\Delta H)^2 + (0.154\,\Delta N)^2}$, ΔH and ΔN is the change of ^{1}H and ^{15}N chemical shifts, respectively, at the current titration point relative to that of the ligand-free state[1]] and standardized into fraction bound value as d/d_{max}. d_{max} is the maximal possible *d* value for the given residue determined by fitting *d* values of that residue at different titration points to the total ligand concentrations with a logistic function which is widely used to model dose-response curves.[2–4]

2) With user-provided values of $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_N and K_C, the predicted fraction bound values of each domain could be determined by eqn (5-11).

3) The values of K_N and K_C were then iteratively refined by a Newton optimization algorithm to match the experimental data, such that the following target function (Sum of squared errors between the experimentally determined data and the predicted data) is minimized globally (insensitive to the starting values):

$$\sum_i \left(\sum_j \left(\frac{d_{i,j}}{d_{max,j}} - f_{N,i} \right)^2 + \sum_k \left(\frac{d_{i,k}}{d_{max,k}} - f_{C,i} \right)^2 \right) \tag{14}$$

Where $d_{i,j}$ is the peak-center CSP value for the resonance from (N)SH2 residue *j* at titration point *i*; $d_{max,j}$ is the maximal possible peak-center CSP value for the resonance from (N)SH2 residue *j*; $f_{N,i}$ is the predicted value of fraction bound for (N)SH2 at titration point *i*. Parameters with index *k* are similarly defined for (C)SH2 residue *k*.

The fitting procedure was performed using home-written scripts in the statistical programming language, R.[5] For each experiment, the best-fit values of parameters were

determined using peak-center CSP values of the same set of residues (G32, L37, L52, H61, Y73, A74, I99, F106, D175, G184, L192, C205, G210, and S244, which were well resolved in both the ligand-free and bound states in all complexes). Fitting errors of the best-fit values were determined by Monte-Carlo analysis using randomly generated synthetic data from a Gaussian distribution for resonances of each domain at each titration point; the mean and standard deviation of the Gaussian distribution were calculated from the experimental data for resonances from a given SH2 domain at a given titration point. In addition, experimental errors were estimated; for each complex, two experiments were performed and processed independently to obtain the experimental errors for the fitted parameters. The variance from experimental errors is generally larger than that from fitting errors; both variances are combined and reported as the final uncertainty for the fitted parameters (see Fig. S3).

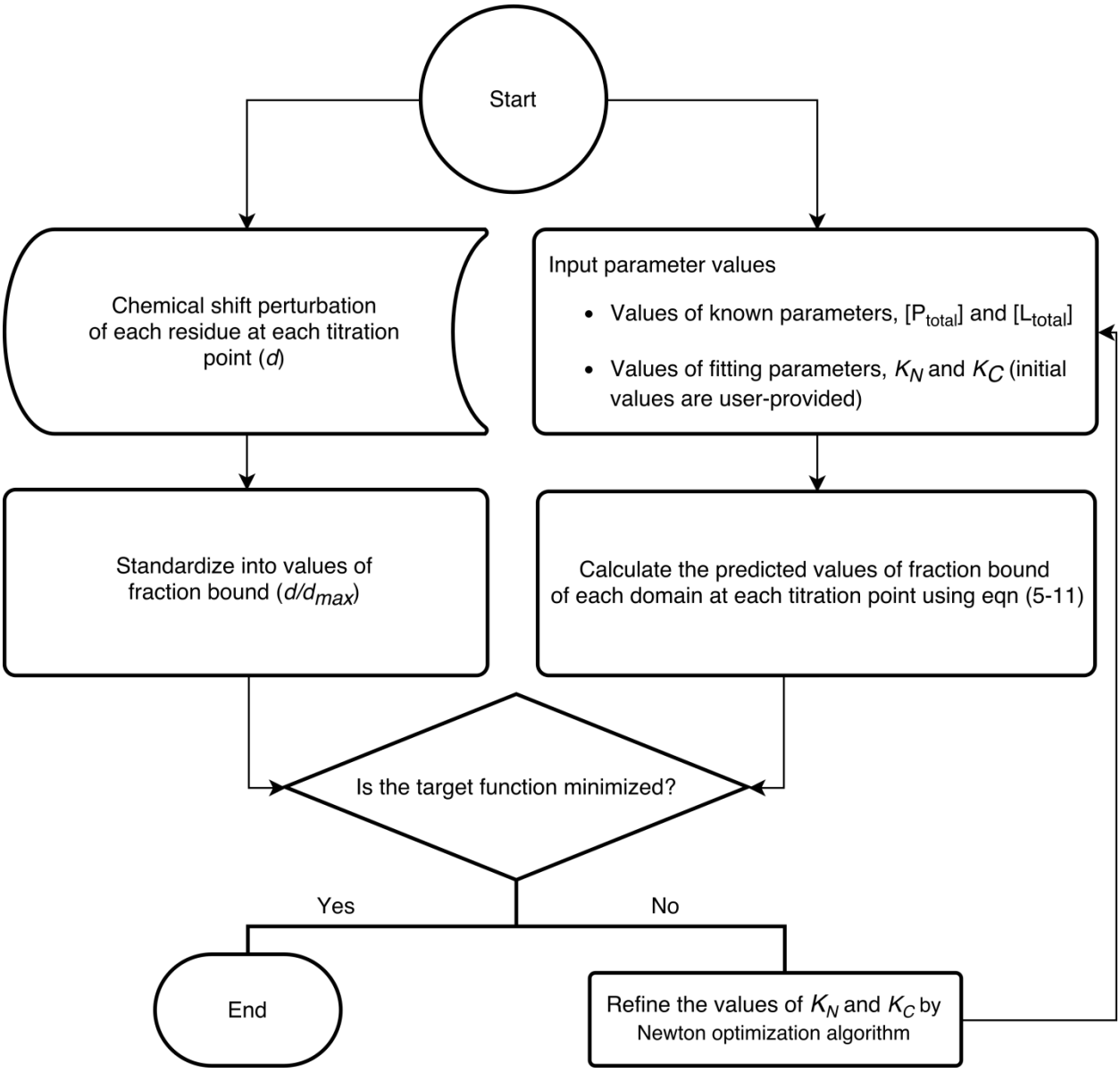

Fig. S2 Flow chart of the titration curve analysis for tSH2/IHP complexes with the 4-state model. The target function to be minimized is the sum of squared errors between the experimentally and computationally determined fraction-bound values of each residue at each titration point.

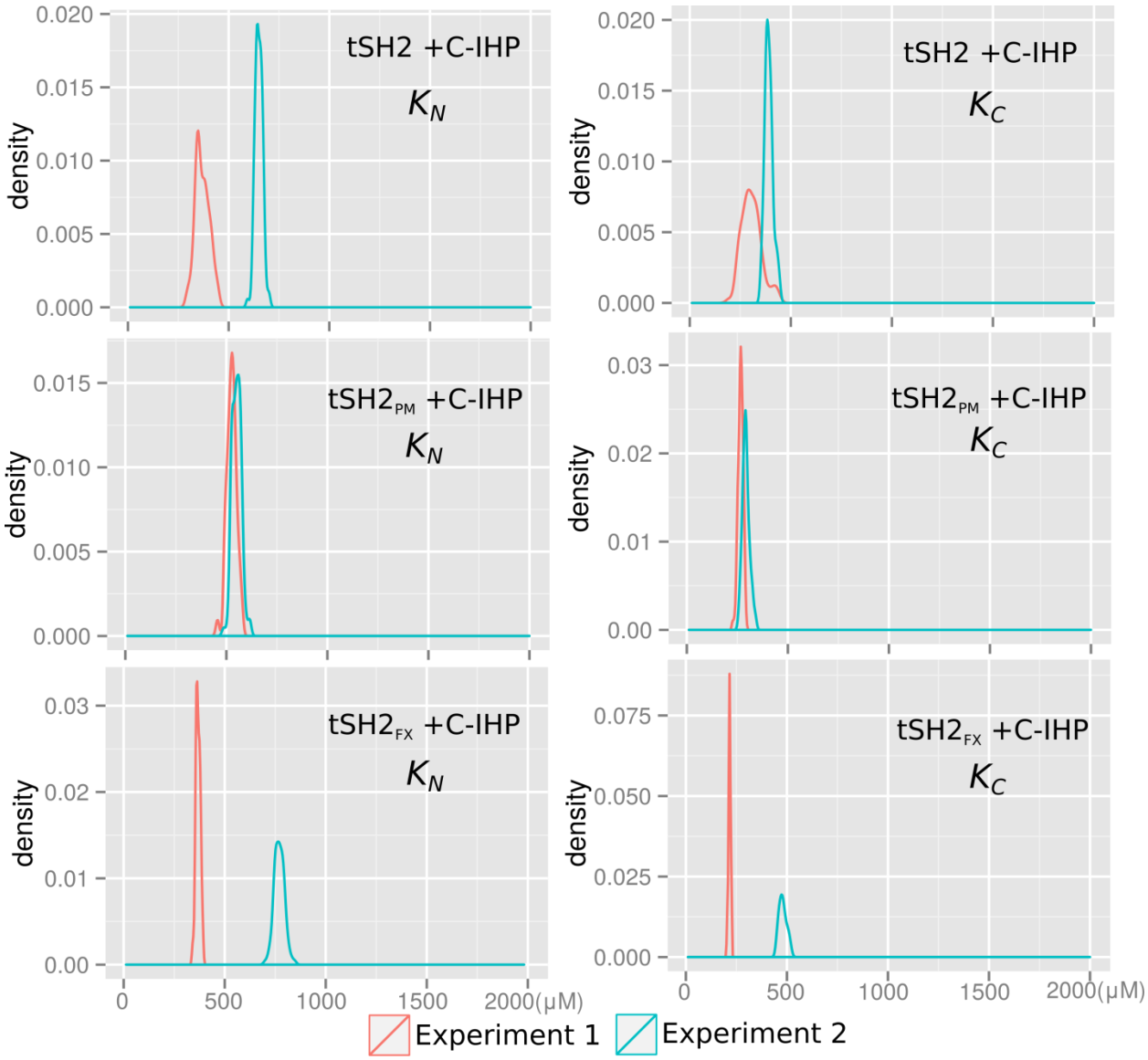

Fig.S3 The distribution of fitted values of dissociation constants of C-IHP binding to the three tSH2 constructs using the **titration-curve analysis** and **4-state model**. Two independent experiments were performed for each complex and 100 Monte-Carlo simulations were performed for each experimental dataset.

2c. Line-shape analysis

The line shape analysis requires two additional parameters to describe the four-state system: the off rate for the (N)SH2 domain binding process k_{off}^{N} and the off rate for the (C)SH2 domain binding process k_{off}^{C}. The corresponding on rates are:

$$k_{on}^{N} = \frac{k_{off}^{N}}{K_N} \tag{15}$$

$$k_{on}^{C} = \frac{k_{off}^{C}}{K_C} \tag{16}$$

The predicted line shape for a given nucleus at a given titration point was generated using the matrix-form solution[6,7] of Bloch-McConnell equations:

$$S^{pred}(\omega) \;=\; \mathrm{Real}\left(\sum((\mathbf{M_1}+\mathbf{M_2})^{-1}\times\mathbf{P})\right) \tag{17}$$

where

$$\mathbf{M_1} = \mathbf{R} - \mathrm{i}\boldsymbol{\Omega} - \mathbf{K} \tag{18}$$

$$\mathbf{M_2} = \mathrm{i}\omega\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \tag{19}$$

$$\mathbf{P} = \begin{pmatrix}[\mathrm{NC}]\\[\mathrm{N^LC}]\\[\mathrm{NC^L}]\\[\mathrm{N^LC^L}]\end{pmatrix} \tag{20}$$

P is a vector of equilibrium concentrations of the four species calculated by eqn (5-9) at the given titration point. **R** is a diagonal matrix with elements equal to the transverse relaxation rates of the resonance of the nucleus in the four species in **P**. **Ω** is a diagonal matrix for the intrinsic resonant frequency (ω_0) of the given nucleus in each of the four species in **P**. **R** and **Ω** were obtained by fitting the resonance peaks at the ligand-free and saturated states of the given nucleus to a Lorentzian function (assuming the peak centers at ω_0 with a line width $2R_{2,0}$). **K** is the 4 by 4 exchange matrix

$$\mathbf{K}=\begin{pmatrix}-(k_{on}^{N}+k_{on}^{C})[\mathrm{L}] & k_{off}^{N} & k_{off}^{C} & 0\\ k_{on}^{N}[\mathrm{L}] & -k_{off}^{N}-k_{on}^{C}[\mathrm{L}] & 0 & k_{off}^{C}\\ k_{on}^{C}[\mathrm{L}] & 0 & -k_{off}^{C}-k_{on}^{N}[\mathrm{L}] & k_{off}^{N}\\ 0 & k_{on}^{C}[\mathrm{L}] & k_{on}^{N}[\mathrm{L}] & -k_{off}^{C}-k_{off}^{N}\end{pmatrix} \tag{21}$$

The line-shape analysis was performed with a procedure (Fig.S4) similar to the titration-curve analysis. In this procedure:

1) The experimental line-shape data for a nucleus at each titration point were extracted from the 2D ^{1}H-^{15}N HSQC spectra as a 1D slice and normalized.
2) With user-provided values of $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_N, K_C, and the off-rates (k_{off}^{N} and k_{off}^{C}), the predicted line-shape data for each residue at each titration point could be determined by eqn (17-21).

3) The values of K_N, K_C and the off-rates were then iteratively refined by a Newton optimization algorithm to match the experimental data, such that the following target function (Sum of squared errors between the experimentally determined and the predicted line-shape data) is minimized globally (insensitive to the starting values):

$$\sum_{i,j,k}\left\{\left[S_{i,j}^{pred}(\omega_k)a_i + b_i\right] - S_{i,j}^{expt}(\omega_k)\right\}^2 \quad (22)$$

Where $S_{i,j}^{pred}(\omega_k)$ and $S_{i,j}^{expt}(\omega_k)$ are the predicted and experimental spectrum intensity of resonance *i* at titration point *j* at frequency ω_k, respectively; a_i and b_i are the intensity and baseline correction factors for resonance *i* to compensate the normalization errors, if any, introduced during the normalization process of the experimental line-shape data.

The line shape-fitting procedure was performed using Matlab R2014a (The MathWorks, Inc.). The experimental 1D-slice data extraction, normalization, and line width/frequency estimation processes utilized tools and subroutines in the IDAP package[8] (version 1.5.4, http://lineshapekin.net). The model-specific calculations of the equilibrium concentrations and the predicted line-shape data were carried out using home-written scripts in Matlab. For each experiment, the best-fit values of parameters were determined using line-shape data of ^{1}H-^{15}N resonances from the same set of residues: G32, L37, L52, H61, Y73, A74, I99, F106, D175, G184, L192, C205, G210, and S244. These resonances were selected because they were well resolved in both the ligand-free and bound states in all complexes. Fitting errors of the best-fit values were determined by bootstrapping analysis using random selections from the experimental datasets for resonances from a given SH2 domain: seven resonances from (N)SH2 and five resonances from (C)SH2. For each complex, two experiments were performed and processed independently to obtain the experimental errors for the fitted parameters. The variance from experimental errors is generally larger than that from fitting errors; both errors are combined and reported as the final uncertainty for the fitted parameters (see Fig. S5).

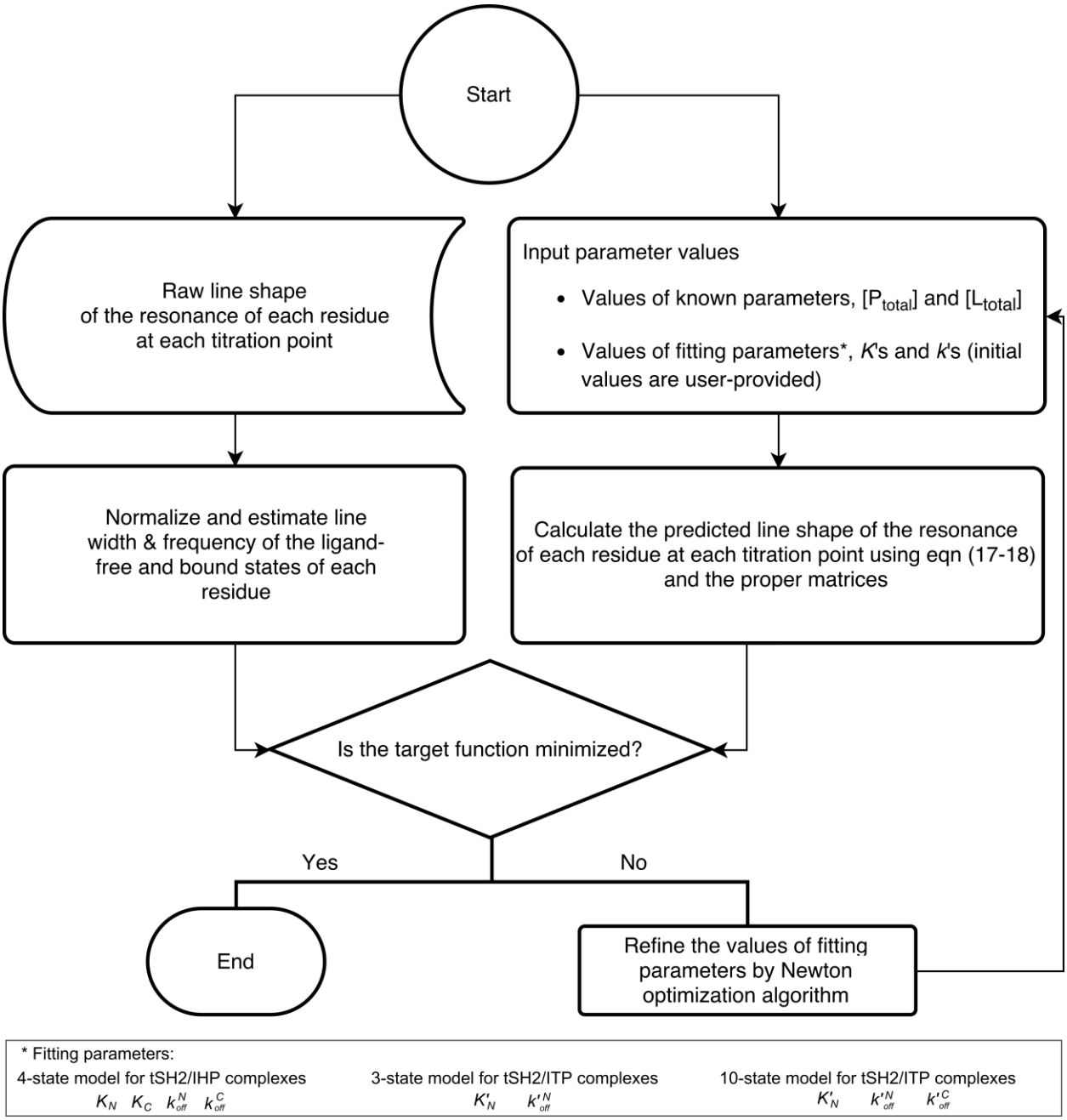

Fig.S4 Flow chart of the line-shape analysis procedures. The target function to be minimized is the sum of squared errors between the experimentally and computationally determined line-shape data of each residue at each titration point. The fitting parameters vary for different models, while the other parameters in the models are fixed, where appropriate, using known values.

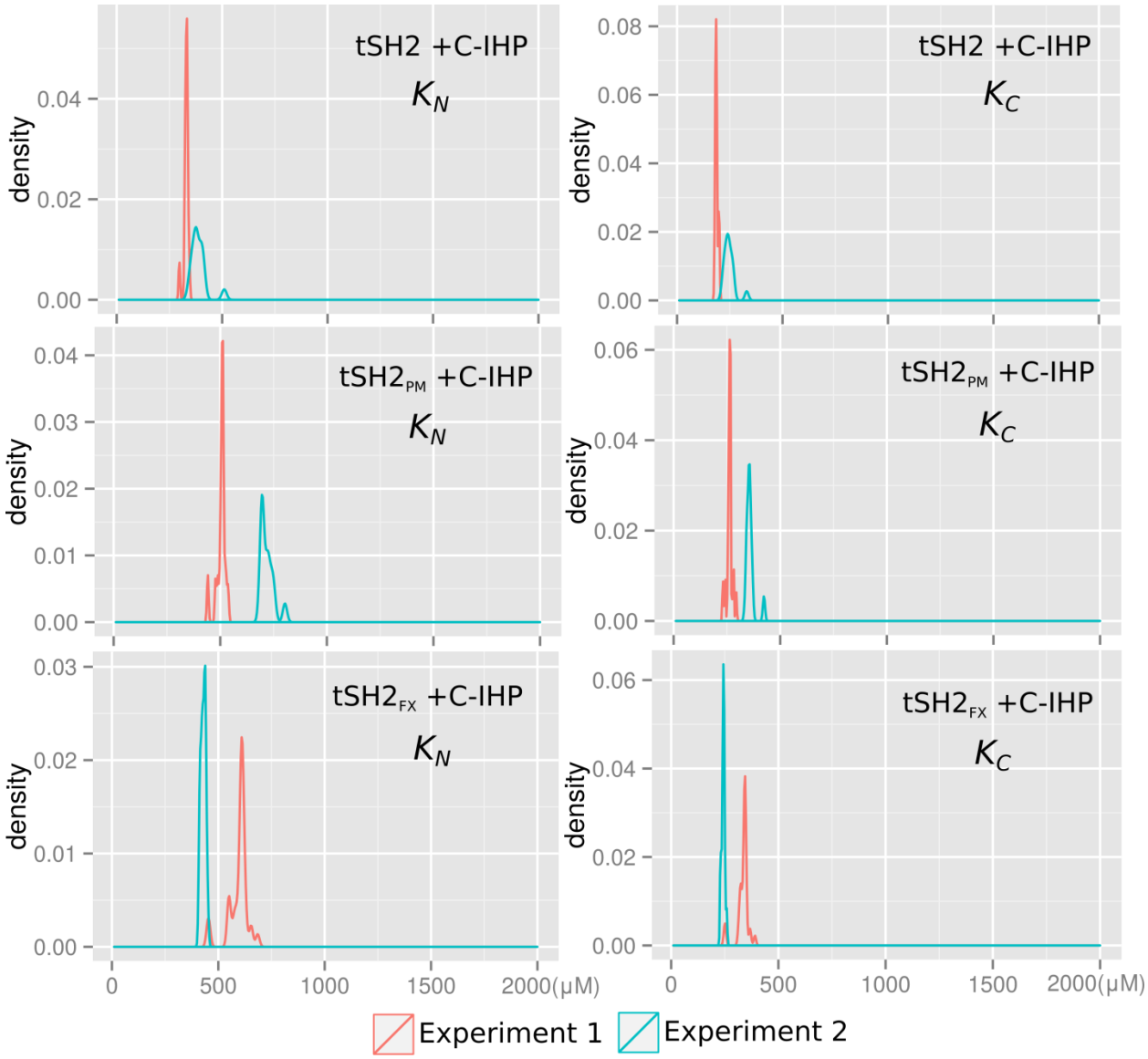

Fig.S5 The distribution of fitted values of dissociation constants of C-IHP binding to the three tSH2 constructs using the **line-shape analysis** and **4-state model**. Two independent experiments were performed for each complex and 100 bootstrapping simulations were performed for each experimental dataset.

2d. The simulated change of concentration of each species

The change of concentration of each species for the tSH2 constructs binding with IHPs in the 4-state binding model were simulated and shown in Fig. S6. Because the binding affinity of a specific SH2 domain with a specific IHP is similar in all three tSH2 constructs (see Table 2 in main text), the simulated curves are similar for all three tSH2 constructs. These curves are described in the Results section of the published text.

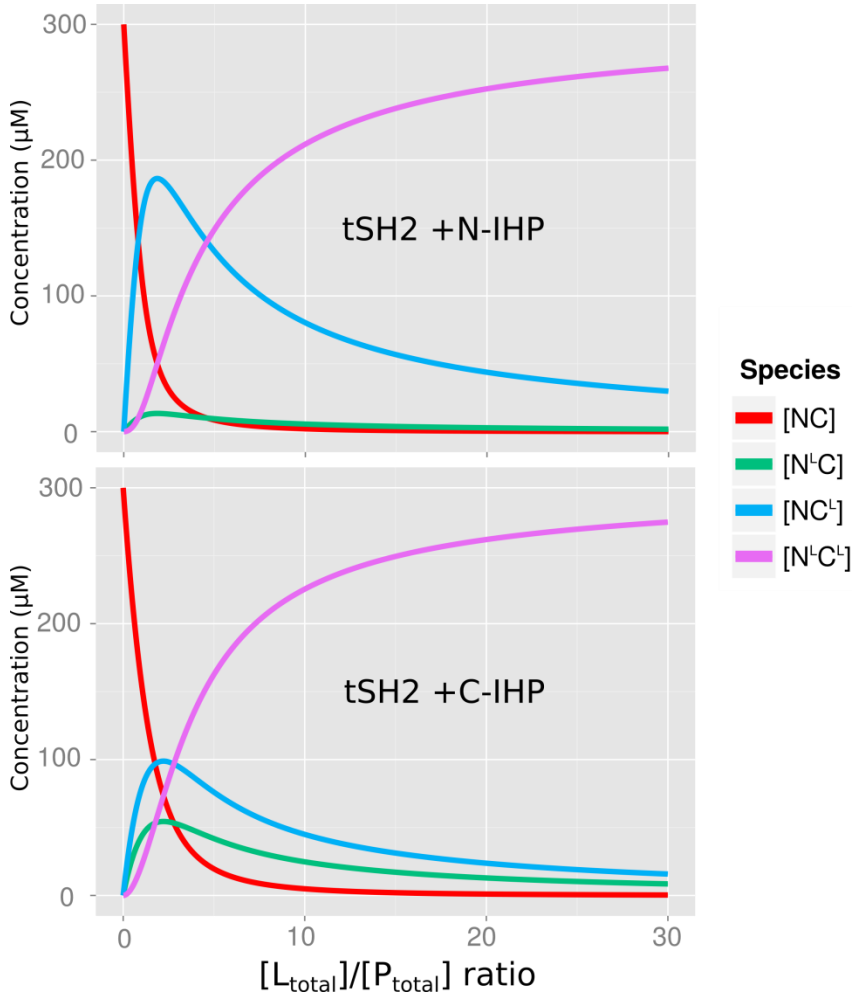

Fig.S6 The simulated change of concentration of each species for the tSH2 construct binding with IHPs in the 4-state binding model. $[P_{total}]$ is 0.3 mM similar as that for NMR experiments. The simulated curves are similar for all three tSH2 constructs, because the binding affinity of a specific SH2 domain with a specific IHP is similar in all three tSH2 constructs. For a specific IHP, the curves shown were simulated using the averaged values of the dissociation constants (K_N and K_C) of the three tSH2 constructs (see Table 2 in main text; K_N and K_C values of bindings with N-IHP or C-IHP by line-shape analysis).

3. The 3-state modeling for tSH2 / ITP complexes

3a. Mathematical descriptions

Considering only the major species, the interaction between the two domains of a tSH2 molecule and the full-length ITP ligand (dp-ITAM peptide, which contains 2 pYs) is described by the scheme in Fig.5b (see main text).

There are only two binding processes: an inter-molecular binding step between the (C)SH2 domain of tSH2 and N-terminal pY cassette of ITP described by K_{CN}, and an intra-molecular binding (isomerization) step between the (N)SH2 domain of tSH2 and C-terminal pY cassette of ITP described by the equilibrium isomerization constant K'_N :

$$K_{CN} = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{NC^N}]} \tag{23}$$

$$K'_N = \frac{[\mathrm{NC^N}]}{[\mathrm{N'C'}]} \tag{24}$$

The total concentration of protein and ligand of each titration point is known and can be expressed as

$$[\mathrm{P_{total}}] = [\mathrm{NC}] + [\mathrm{NC^N}] + [\mathrm{N'C'}] \tag{25}$$

$$[\mathrm{L_{total}}] = [\mathrm{L}] + [\mathrm{NC^N}] + [\mathrm{N'C'}] \tag{26}$$

From eqn (23-26), the relationship between [L], [$\mathrm{P_{total}}$], [$\mathrm{L_{total}}$], K_{CN}, and K'_N is

$$\begin{aligned}(K'_N + 1)[\mathrm{L}]^2 + \big((K'_N + 1)[\mathrm{P_{total}}] - (K'_N + 1)[\mathrm{L_{total}}] + K_{CN}K'_N\big)[\mathrm{L}] \\ - K_{CN}K'_N[\mathrm{L_{total}}] = 0\end{aligned} \tag{27}$$

With given values of [$\mathrm{P_{total}}$], [$\mathrm{L_{total}}$], K_{CN}, K'_N, the value of [L] can be solved analytically. Once [L] is known, the concentrations of other species in the system could be determined as follows based on eqn (23-26):

$$[\mathrm{NC}] = \frac{K_{CN}K'_N[\mathrm{P_{total}}]}{(K'_N + 1)[\mathrm{L}] + K_{CN}K'_N} \tag{28}$$

$$[\mathrm{NC^N}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_{CN}} \tag{29}$$

$$[\mathrm{N'C'}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K'_N K_{CN}} \tag{30}$$

3b. The line-shape analysis

The line-shape analysis requires two more parameters to describe the 3-state system: the off rates for the (C)SH2 domain binding process, k_{off}^{CN}, and that for the(N)SH2 domain binding process k'^{N}_{off}. The corresponding on rates are:

$$k_{on}^{CN} = \frac{k_{off}^{CN}}{K_{CN}} \tag{31}$$

$$k'^{N}_{on} = \frac{k'^{N}_{off}}{K'_{N}} \tag{32}$$

The predicted line shape for a nucleus at a given titration point using the 3-state model is as described in section 2c, except that the following matrices in eqn (17-18) are now:

$$\mathbf{M_2} = \mathrm{i}\omega \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{33}$$

$$\mathbf{P} = \begin{pmatrix} [\mathrm{NC}] \\ [\mathrm{NC^N}] \\ [\mathrm{N'C'}] \end{pmatrix} \tag{34}$$

P is a vector of equilibrium concentrations of the three species calculated with eqn (27-30) at the given titration point. **R** and **Ω** were obtained as described for the 4-state model. **K** is the 3 by 3 exchange matrix

$$\mathbf{K} = \begin{pmatrix} -k^{CN}_{on}[\mathrm{L}] & k^{CN}_{off} & 0 \\ k^{CN}_{on}[\mathrm{L}] & -k^{CN}_{off} - k'^{N}_{on} & k'^{N}_{off} \\ 0 & k'^{N}_{on} & -k'^{N}_{off} \end{pmatrix} \tag{35}$$

3c. The simulated change of concentration of each species

The change of concentration of each species for the tSH2 constructs binding with ITP in the 3-state binding model were simulated and shown in Fig. S7. These curves are discussed in the Results section of the published text.

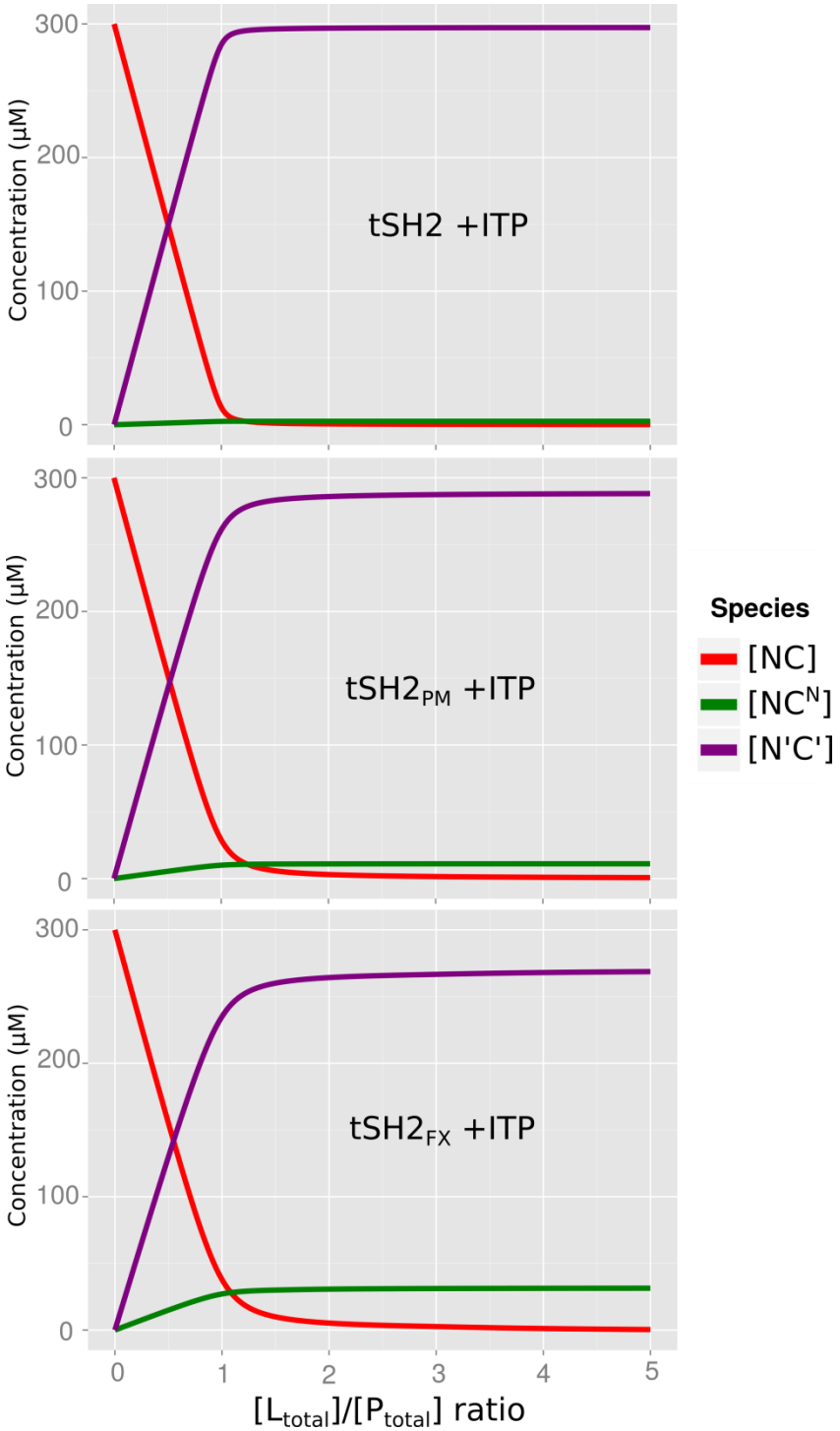

Fig.S7 The simulated change of concentration of each species for the three tSH2 constructs binding with ITP in the 3-state binding model. $[P_{total}]$ is 0.3 mM similar as that for NMR experiments. For each complex, the curves were simulated using the fitted intra-molecular isomerization constant (K'_N) from the 3-state model (see Table 3 in main text) and the corresponding inter-molecular dissociation constant (K_{CN}) from the N-IHP complexes (see Table 2 in main text; the K_C values of bindings with N-IHP by line-shape analysis).

4. The 10-state modeling for tSH2 / ITP complexes

4a. Mathematical descriptions

Considering all possible species, the interaction between the two domains of a tSH2 molecule and an ITP ligand (dp-ITAM peptide, which contains 2 pYs) is described by the scheme in Fig.5a (see main text).

Based on this scheme, the equilibria for the initial inter-molecular binding processes (a single tSH2 domain binding with a single pY of an ITP ligand) are characterized by the following equations:

$$K_{NN} = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{N^N C}]} = \frac{[\mathrm{NC^N}][\mathrm{L}]}{[\mathrm{N^N C^N}]} = \frac{[\mathrm{NC^C}][\mathrm{L}]}{[\mathrm{N^N C^C}]} \tag{36}$$

$$K_{NC} = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{N^C C}]} = \frac{[\mathrm{NC^N}][\mathrm{L}]}{[\mathrm{N^C C^N}]} = \frac{[\mathrm{NC^C}][\mathrm{L}]}{[\mathrm{N^C C^C}]} \tag{37}$$

$$K_{CN} = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{NC^N}]} = \frac{[\mathrm{N^N C}][\mathrm{L}]}{[\mathrm{N^N C^N}]} = \frac{[\mathrm{N^C C}][\mathrm{L}]}{[\mathrm{N^C C^N}]} \tag{38}$$

$$K_{CC} = \frac{[\mathrm{NC}][\mathrm{L}]}{[\mathrm{NC^C}]} = \frac{[\mathrm{N^N C}][\mathrm{L}]}{[\mathrm{N^N C^C}]} = \frac{[\mathrm{N^C C}][\mathrm{L}]}{[\mathrm{N^C C^C}]} \tag{39}$$

The equilibria for the two intra-molecular binding processes (from the 1:1 mono-functional complex $\mathrm{N^C C}$ or $\mathrm{NC^N}$ to form a 1:1 bi-functional complex N′C′) are characterized by the equilibrium isomerization constants:

$$K'_N = \frac{[\mathrm{NC^N}]}{[\mathrm{N'C'}]} \tag{40}$$

$$K'_C = \frac{[\mathrm{N^C C}]}{[\mathrm{N'C'}]} \tag{41}$$

Note that K'_C can be expressed as $K_{CN} \times K'_N / K_{NC}$ and is thus not an independent variable. The total concentration of protein and ligand of each titration point is known and can be expressed as

$$\begin{aligned}[\mathrm{P_{total}}] = {} & [\mathrm{NC}] + [\mathrm{N^N C}] + [\mathrm{N^C C}] + [\mathrm{NC^N}] + [\mathrm{NC^C}] + [\mathrm{N'C'}] \\ & + [\mathrm{N^N C^N}] + [\mathrm{N^C C^N}] + [\mathrm{N^N C^C}] + [\mathrm{N^C C^C}]\end{aligned} \tag{42}$$

$$\begin{aligned}[\mathrm{L_{total}}] = {} & [\mathrm{L}] + [\mathrm{N^N C}] + [\mathrm{N^C C}] + [\mathrm{NC^N}] + [\mathrm{NC^C}] + [\mathrm{N'C'}] \\ & + 2[\mathrm{N^N C^N}] + 2[\mathrm{N^C C^N}] + 2[\mathrm{N^N C^C}] + 2[\mathrm{N^C C^C}]\end{aligned} \tag{43}$$

From eqn (36-43), the relationship between [L], $[\mathrm{P_{total}}]$, $[\mathrm{L_{total}}]$, K_{NN}, K_{NC}, K_{CN}, K_{CC}, and K'_N is

$$a[\mathrm{L}]^3 + b[\mathrm{L}]^2 + c[\mathrm{L}] + d = 0 \tag{44}$$

where

$$a = (K'_N K_{CN} + K'_N K_{CC}) K_{NN} + (K'_N K_{CN} + K'_N K_{CC}) K_{NC} \tag{45}$$

$$\begin{aligned} b = {} & \big((2\,K'_N K_{CN} + 2\,K'_N K_{CC}) K_{NN} + (2\,K'_N K_{CN} + 2\,K'_N K_{CC}) K_{NC}\big)[\mathrm{P_{total}}] \\ & + \big((-K'_N K_{CN} - K'_N K_{CC}) K_{NN} + (-K'_N K_{CN} - K'_N K_{CC}) K_{NC}\big)[\mathrm{L_{total}}] \\ & + \big((K'_N K_{CN} + (K'_N + 1) K_{CC}) K_{NC} + K'_N K_{CC} K_{CN}\big) K_{NN} \\ & + K'_N K_{CC} K_{CN} K_{NC} \end{aligned} \tag{46}$$

$$c = \Big(\big((K'_N K_{CN} + (K'_N + 1)K_{CC})K_{NC} + K'_N K_{CC} K_{CN}\big)K_{NN} + K'_N K_{CC} K_{CN} K_{NC}\Big)[\mathrm{P_{total}}] + \Big(\Big(\big((-K'_N - 1)K_{CC} - K'_N K_{CN}\big)K_{NC} - K'_N K_{CC} K_{CN}\Big)K_{NN} - K'_N K_{CC} K_{CN} K_{NC}\Big)[\mathrm{L_{total}}] + K'_N K_{CC} K_{CN} K_{NC} K_{NN} \quad (47)$$

$$d = -K'_N K_{CC} K_{CN} K_{NC} K_{NN}[\mathrm{L_{total}}] \quad (48)$$

With given values of $[P_{total}]$, $[L_{total}]$, K_{NN}, K_{NC}, K_{CN}, K_{CC}, and K'_N, the value of [L] can be solved analytically (by Maxima). Once [L] is known, the concentrations of other species in the system are as follows based on eqn (36-43):

$$[\mathrm{NC}] = K'_N K_{CC} K_{CN} K_{NC} K_{NN}[\mathrm{P_{total}}] \Big/ \Big(\big((K'_N K_{CN} + K'_N K_{CC})K_{NN} + (K'_N K_{CN} + K'_N K_{CC})K_{NC}\big)[\mathrm{L}]^2 + \Big(\big((K'_N K_{CN} + (K'_N + 1)K_{CC})K_{NC} + K'_N K_{CC} K_{CN}\big)K_{NN} + K'_N K_{CC} K_{CN} K_{NC}\Big)[\mathrm{L}] + K'_N K_{CC} K_{CN} K_{NC} K_{NN}\Big) \quad (49)$$

$$[\mathrm{N^N C}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_{NN}} \quad (50)$$

$$[\mathrm{N^C C}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_{NC}} \quad (51)$$

$$[\mathrm{NC^N}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_{CN}} \quad (52)$$

$$[\mathrm{NC^C}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K_{CC}} \quad (53)$$

$$[\mathrm{N'C'}] = \frac{[\mathrm{L}][\mathrm{NC}]}{K'_N K_{CN}} \quad (54)$$

$$[\mathrm{N^N C^N}] = \frac{[\mathrm{L}]^2[\mathrm{NC}]}{K_{CN} K_{NN}} \quad (55)$$

$$[\mathrm{N^C C^N}] = \frac{[\mathrm{L}]^2[\mathrm{NC}]}{K_{CN} K_{NC}} \quad (56)$$

$$[\mathrm{N^N C^C}] = \frac{[\mathrm{L}]^2[\mathrm{NC}]}{K_{CC} K_{NN}} \quad (57)$$

$$[\mathrm{N^C C^C}] = \frac{[\mathrm{L}]^2[\mathrm{NC}]}{K_{CC} K_{NC}} \quad (58)$$

4b. The line-shape analysis

The line-shape analysis requires six more parameters to describe the 10-state system: the off rates for the four inter-molecular and two intra-molecular binding processes. The corresponding

on rates are calculated as follows:

$$k_{on}^{NN} = \frac{k_{off}^{NN}}{K_{NN}} \tag{59}$$

$$k_{on}^{NC} = \frac{k_{off}^{NC}}{K_{NC}} \tag{60}$$

$$k_{on}^{CN} = \frac{k_{off}^{CN}}{K_{CN}} \tag{61}$$

$$k_{on}^{CC} = \frac{k_{off}^{CC}}{K_{CC}} \tag{62}$$

$$k'^{N}_{on} = \frac{k'^{N}_{off}}{K'_{N}} \tag{63}$$

$$k'^{C}_{on} = \frac{k'^{C}_{off}}{K'_{C}} \tag{64}$$

The predicted line shape for a nucleus at a given titration point using the 10-state model are as described for the 4-state model in section 2c, except that the matrices in eqn (17-18) become:

$$\mathbf{M_2} = \mathrm{i}\omega\mathbf{E} \tag{65}$$

$$\mathbf{P} = \begin{pmatrix} [\mathrm{NC}] \\ [\mathrm{N^N C}] \\ [\mathrm{N^C C}] \\ [\mathrm{NC^N}] \\ [\mathrm{NC^C}] \\ [\mathrm{N'C'}] \\ [\mathrm{N^N C^N}] \\ [\mathrm{N^C C^N}] \\ [\mathrm{N^N C^C}] \\ [\mathrm{N^C C^C}] \end{pmatrix} \tag{66}$$

E is a 10 by 10 identity matrix. **P** is a vector of equilibrium concentrations of the ten species calculated from eqn (44-58) for the given titration point. **R** and **Ω** were obtained as described for the 4-state model. For a given nucleus, binding with the N-terminal pY of ITP is assumed to result in similar values of ω_0 and $R_{2,0}$ in the bound state as binding with the C-terminal pY, because they showed similar titration profiles in terms of the direction and magnitude of CSP (chemical shift perturbation). **K** is the 10 by 10 exchange matrix

$$\mathbf{K}=\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & 0 & 0 & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & 0 & 0 & a_{2,7} & 0 & a_{2,9} & 0 \\ a_{3,1} & 0 & a_{3,3} & 0 & 0 & a_{3,6} & 0 & a_{3,8} & 0 & a_{3,10} \\ a_{4,1} & 0 & 0 & a_{4,4} & 0 & a_{4,6} & a_{4,7} & a_{4,8} & 0 & 0 \\ a_{5,1} & 0 & 0 & 0 & a_{5,5} & 0 & 0 & 0 & a_{5,9} & a_{5,10} \\ 0 & 0 & a_{6,3} & a_{6,4} & 0 & a_{6,6} & 0 & 0 & 0 & 0 \\ 0 & a_{7,2} & 0 & a_{7,4} & 0 & 0 & a_{7,7} & 0 & 0 & 0 \\ 0 & 0 & a_{8,3} & a_{8,4} & 0 & 0 & 0 & a_{8,8} & 0 & 0 \\ 0 & a_{9,2} & 0 & 0 & a_{9,5} & 0 & 0 & 0 & a_{9,9} & 0 \\ 0 & 0 & a_{10,3} & 0 & a_{10,5} & 0 & 0 & 0 & 0 & a_{10,10} \end{pmatrix} \quad (67)$$

Where

$$a_{1,1} = -(k_{on}^{NN} + k_{on}^{NC} + k_{on}^{CN} + k_{on}^{CC})[\mathrm{L}] \quad (68)$$

$$a_{2,1} = k_{on}^{NN}[\mathrm{L}] \quad (69)$$

$$a_{3,1} = k_{on}^{NC}[\mathrm{L}] \quad (70)$$

$$a_{4,1} = k_{on}^{CN}[\mathrm{L}] \quad (71)$$

$$a_{5,1} = k_{on}^{CC}[\mathrm{L}] \quad (72)$$

$$a_{1,2} = k_{off}^{NN} \quad (73)$$

$$a_{2,2} = -k_{off}^{NN} - (k_{on}^{CN} + k_{on}^{CC})[\mathrm{L}] \quad (74)$$

$$a_{7,2} = k_{on}^{CN}[\mathrm{L}] \quad (75)$$

$$a_{9,2} = k_{on}^{CC}[\mathrm{L}] \quad (76)$$

$$a_{1,3} = k_{off}^{NC} \quad (77)$$

$$a_{3,3} = -k_{off}^{NC} - k'^{C}_{on} - (k_{on}^{CN} + k_{on}^{CC})[\mathrm{L}] \quad (78)$$

$$a_{6,3} = k'^{C}_{on} \quad (79)$$

$$a_{8,3} = k_{on}^{CN}[\mathrm{L}] \quad (80)$$

$$a_{10,3} = k_{on}^{CC}[\mathrm{L}] \quad (81)$$

$$a_{1,4} = k_{off}^{CN} \quad (82)$$

$$a_{4,4} = -k_{off}^{CN} - k'^{N}_{on} - (k_{on}^{NN} + k_{on}^{NC})[\mathrm{L}] \quad (83)$$

$$a_{6,4} = k'^{N}_{on} \quad (84)$$

$$a_{7,4} = k_{on}^{NN}[\mathrm{L}] \quad (85)$$

$$a_{8,4} = k_{on}^{NC}[\mathrm{L}] \quad (86)$$

$$a_{1,5} = k_{off}^{CC} \quad (87)$$

$$a_{5,5} = -k_{off}^{CC} - (k_{on}^{NN} + k_{on}^{NC})[\mathrm{L}] \quad (88)$$

$$a_{9,5} = k_{on}^{NN}[\mathrm{L}] \quad (89)$$

$$a_{10,5} = k_{on}^{NC}[\mathrm{L}] \quad (90)$$

$$a_{3,6} = k'^{C}_{off} \quad (91)$$

$$a_{4,6} = k'^{N}_{off} \quad (92)$$

$$a_{6,6} = -k'^{C}_{off} - k'^{N}_{off} \quad (93)$$

$$a_{2,7} = k_{off}^{CN} \quad (94)$$

$$a_{4,7} = k_{off}^{NN} \quad (95)$$

$$a_{7,7} = -k_{off}^{CN} - k_{off}^{NN} \quad (96)$$

$$a_{3,8} = k_{off}^{CN} \quad (97)$$

$$a_{4,8} = k_{off}^{NC} \quad (98)$$

$$a_{8,8} = -k_{off}^{CN} - k_{off}^{NC} \quad (99)$$

$$a_{2,9} = k_{off}^{CC} \quad (100)$$

$$a_{5,9} = k_{off}^{NN} \quad (101)$$

$$a_{9,9} = -k_{off}^{CC} - k_{off}^{NN} \quad (102)$$

$$a_{3,10} = k_{off}^{CC} \quad (103)$$

$$a_{5,10} = k_{off}^{NC} \quad (104)$$

$$a_{10,10} = -k_{off}^{CC} - k_{off}^{NC} \quad (105)$$

4c. The simulated change of concentration of each species

The change of concentration of each species for the tSH2 constructs binding with ITP in the 10-state binding model were simulated and shown in Fig. S8. These curves are described in the Results section of the published text.

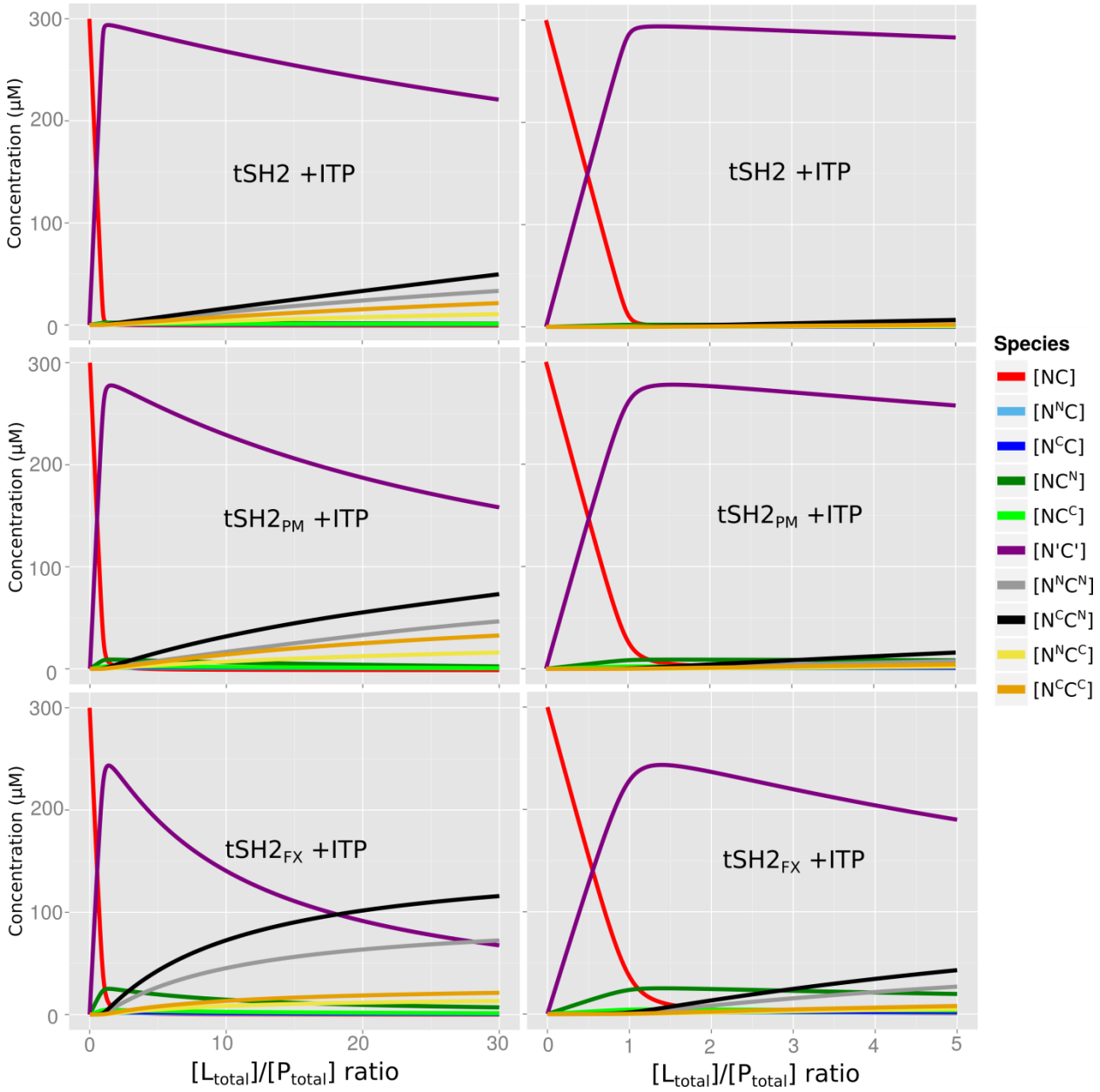

Fig.S8 The simulated change of concentration of each species for the tSH2 constructs binding with ITP in the 10-state binding model, with two ranges of the $[L_{total}]/[P_{total}]$ ratio for each construct. $[P_{total}]$ is 0.3 mM similar as that for NMR experiments. For each construct, the curves were simulated using the fitted intra-molecular isomerization constant (K'_N) from the 10-state model (see Table 3 in main text) and the corresponding inter-molecular dissociation constants (K_{NN}, K_{NC}, K_{CN}, and K_{CC}) from the IHP complexes (see Table 2 in main text; K_N and K_C values of bindings with N-IHP and C-IHP by line-shape analysis).

References

1 A. Ayed, F. A. Mulder, G. S. Yi, Y. Lu, L. E. Kay and C. H. Arrowsmith, *Nat. Struct. Biol.*, 2001, **8**, 756–760.

2 A. DeLean, P. J. Munson and D. Rodbard, *Am. J. Physiol.*, 1978, **235**, E97–102.

3 C. Ritz, *Environ. Toxicol. Chem.*, 2010, **29**, 220–229.

4 C. Ritz and J. C. Streibig, *J. Stat. Softw.*, 2005, **12**, 1–22.

5 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.

6 A. I. Greenwood, M. J. Rogals, S. De, K. P. Lu, E. L. Kovrigin and L. K. Nicholson, *J. Biomol. NMR*, 2011, **51**, 21–34.

7 C. B. Post, *Methods Enzymol.*, 1994, **240**, 438–446.

8 E. L. Kovrigin, *J. Biomol. NMR*, 2012, **53**, 257–270.