PAPER

CrossMark
←click for updates

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

Chao Feng and Carol Beth Post*

The phosphorylation of interdomain A (IA), a linker region between tandem SH2 domains of Syk tyrosine kinase, regulates the binding affinity for association of Syk with doubly-phosphorylated ITAM regions of the B cell receptor. The mechanism of this allosteric regulation has been suggested to be a switch from the high-affinity bifunctional binding, mediated through both SH2 domains binding two phosphotyrosine residues of ITAM, to a substantially lower-affinity binding of only one SH2 domain. IA phosphorylation triggers the switch by inducing disorder in IA and weakening the SH2–SH2 interaction. The postulated switch to a single-SH2-domain binding mode is examined using NMR to monitor site-specific binding to each SH2 domain of Syk variants engineered to have IA regions that differ in conformational flexibility. The combined analysis of titration curves and NMR line-shapes provides sufficient information to determine the energetics of inter-molecular binding at each SH2 site along with an intra-molecular binding or isomerization step. A less favorable isomerization equilibrium associated with the changes in the SH2–SH2 conformational ensemble and IA flexibility accounts for the inhibition of Syk association with membrane ITAM regions when IA is phosphorylated, and refutes the proposed switch to single-SH2-domain binding. Syk localizes in the cell through its SH2 interactions, and this basis for allosteric regulation of ITAM association proposes for the first time a phosphorylation-dependent model to regulate Syk binding to alternate receptors and other signaling proteins that differ either in the number of residues separating ITAM phosphotyrosines or by having only one phosphotyrosine, a half ITAM.

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
† Electronic supplementary information (ESI) available: The mathematical descriptions and fitting procedures of models, error analysis, simulated changes of concentration of each species of each protein/ligand system. See DOI: 10.1039/c5cp05417f

Introduction

The non-receptor spleen tyrosine kinase (Syk) is a central regulatory protein mediating signal transduction from immunoreceptors to various downstream targets[1] and is widely expressed in haematopoietic cells (B cells, mast cells, neutrophiles and macrophages). Syk activity is closely related with many vital cellular events including cell proliferation, differentiation, polarization, cell–cell adhesion and phagocytosis.[2] Given its primary role in signaling, it is clear why dysfunction of Syk has been implicated in numerous inflammatory and autoimmune disorders (*e.g.* allergy, asthma, allergic rhinitis, atopic dermatitis, immune thrombocytopenia purpura, autoimmune hemolytic anemia, rheumatoid arthritis, multiple sclerosis and systemic lupus erythematosus).[3] Abnormal expression of Syk is also associated with human B-cell malignancies and cancers (such as breast cancers, gastric cancers and melanomas),[3–5] although the role of Syk in tumor progression is poorly

understood; Syk appears to suppress tumor formation early in tumor genesis while being a promoter of tumor progression at later stages.[6] The central function in numerous signaling pathways and phosphorylation of multiple protein substrates demands tight regulation of Syk activity. A complete characterization and understanding of the regulatory mechanisms of Syk activity is therefore particularly important given the disease relevance of Syk.

Inflammatory responses initiated by extracellular ligands binding membrane receptors lead to phosphorylation of cytosolic regions of the receptor known as immunoreceptor tyrosine-based activation motifs, ITAMs,[7,8] and recruitment of Syk to the membrane. ITAMs contain two YXX(I/L) cassettes separated by 6–10 amino acids.[9] Both tyrosine residues of ITAM need to be phosphorylated for optimal signaling.[1] In B cells, the Src-family kinase Lyn phosphorylates the N-terminal YXX(I/L) cassette of the B-cell receptor but not the C-terminal one,[1,10] which is likely phosphorylated by Syk itself.[1] Doubly phosphorylated ITAM, dp-ITAM, thus serves as a docking site for Syk *via* its two tandem SH2 domains. The association of Syk with receptor dp-ITAM is high-affinity (nM),[11–16] and negatively regulated by tyrosine phosphorylation of Syk at position 130 (murine numbering), which lowers affinity, releasing Syk from the receptor.

This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 5807–5818 | 5807

The structural basis for a high-affinity interaction of Syk with dp-ITAM is readily explained by the crystallographic structure. Syk tyrosine kinase comprises two SH2 domains, (N)SH2 and (C)SH2 connected by interdomain A (IA), followed by interdomain B (IB) and a catalytic domain (Fig. 1a). The two SH2 domains together with IA are known as tandem SH2 domains (tSH2) and have a high sequence identity between human and murine Syk (92% for (N)SH2, 94% for (C)SH2, and 100% for IA). In the crystal form of the tSH2 fragment in complex with a dp-ITAM peptide, the SH2 domains are oriented with the two pY pockets optimally positioned to fit the spacing of the pYXX(I/L) cassettes of dp-ITAM in a head-to-tail fashion[17] (Fig. 1b). This optimal spacing of the phosphotyrosine pockets exists in the six molecules in the asymmetric unit of the crystal structure (PDB ID 1A81), which differ in relative orientation of the two SH2 domains by <18°. High-affinity binding thus derives from the large contact surface and combined association

of tSH2 with two pYXX(I/L) cassettes. Interestingly, we note previously reported NMR residual dipolar coupling data established that in solution the SH2 domain–domain orientation of the unbound tSH2 resembles the orientation in the crystal structure of the dp-ITAM-bound tSH2, not unbound tSH2,[18] which differ in orientation by ~50°. The unbound SH2–SH2 domain structure is similar in the crystal structure for tSH2[19] and full-length Syk,[20] so that neither have SH2–SH2 orientations consistent with the NMR residual dipolar couplings.

The regulatory tyrosine, Y130 (murine numbering; Y131 in human), lies in IA, distant to the phosphotyrosine binding pockets of the SH2 domains (Fig. 1b). The effect of phosphorylation of Y130 in IA is to dissociate Syk from ITAM receptors, freeing Syk to engage in cytosolic or nuclear signal transduction pathways.[21–23] The introduction of negative charge at position 130 in Syk IA reduces the binding affinity between Syk tSH2 and dp-ITAM peptides by 2–3 orders of magnitude, and in the range of binding affinities that correspond to a single SH2 domain binding to phosphotyrosine peptides.[13,14,18] Because Y130 is distal to the binding pockets of the SH2 domains (>20 angstrom; see Fig. 1b) with no evidence for competition with the pYXX(I/L) cassettes for binding, this regulation is an allosteric control mechanism. It is distinguished from most other phosphorylation regulatory mechanisms in which tyrosine phosphorylation controls protein–protein interaction by direct effects of phosphorylation, either promoting binding by generation of the recognition determinant or inhibiting binding by steric conflict.[24]

Earlier studies investigating the effects of Y130 phosphorylation on dynamics and conformation of Syk tandem SH2 showed that negative charge introduced into IA at position 130 causes large changes in NMR relaxation and sedimentation velocity, indicating that phosphorylation leads to a change in the conformational ensemble causing reorientation in the relative SH2 domain structure and a substantial weakening of the SH2–SH2 interaction.[18] The conformational effect of phosphorylation is of considerably greater magnitude than the flexibility previously described for unphosphorylated Syk.[16] It was suggested that the allosteric inhibition of the Syk–receptor interaction by IA phosphorylation was through a conformational change in domain structure that switched the bifunctional high-affinity binding across two SH2 domains to a monofunctional lower-affinity binding that involved only one SH2 domain.[18] Thus, linker A phosphorylation was proposed to restrict the separation and relative orientation of two SH2 domains in a manner that no longer appropriately positioned the two phosphotyrosine binding sites to allow bifunctional binding, and therefore binding is by a single pYXX(I/L) cassette and the affinity for ITAM is reduced to a value comparable to a single SH2 domain.

Here we explore the molecular basis of the allosteric mechanism of IA phosphorylation in regulating Syk association with receptor ITAMs by affecting domain–domain structure. The consequence of Y130 phosphorylation on the ensemble of SH2–SH2 domain conformations was monitored by NMR relaxation. In addition, we exploited the residue-level information of NMR chemical shift to decompose the contribution to binding from each SH2 domain using a multistate analysis of binding equilibrium. We find that even though negative charge in IA does not completely
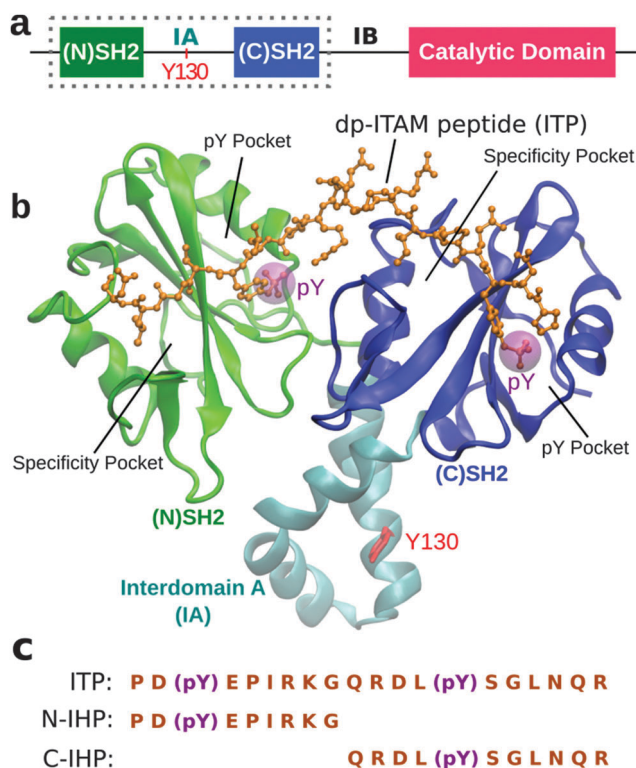

Fig. 1 The structure of Syk tyrosine kinase. (a) The domain structure of Syk: (N)SH2 domain (8–118), (C)SH2 domain (163–264), and catalytic domain (364–620) are connected by interdomain A (IA, 119–162) and interdomain B (IB, 265–363). The tandem SH2 domains (tSH2) (dotted box) contain (N)SH2, IA, and (C)SH2 (residues are numbered according to murine Syk). (b) The ribbon representation of Syk tSH2 bound with a doubly phosphorylated ITAM (dp-ITAM) peptide (balls and sticks in orange) (PDB ID 1A81 chain A and B). The six molecules in the asymmetric unit of the crystal structure bind dp-ITAM peptide with the same spacing of pY and differ in relative orientation of the two SH2 domains by <18°. Green: (N)SH2. Cyan: interdomain A. Blue: (C)SH2. Y130 in IA is shown with red sticks. The dp-ITAM peptide (ITP) is derived from the CD3ε chain of the T cell receptor. The phosphoryl groups of the two phosphotyrosines (pYs) on ITP are shown as purple spheres. Binding is head-to-tail: (N)SH2 binds the C-terminal pY and (C)SH2 binds the N-terminal pY. (c) The sequence of ITP and the two half peptides, N-IHP and C-IHP.

5808 | *Phys. Chem. Chem. Phys.*, 2016, **18**, 5807–5818

This journal is © the Owner Societies 2016

disorder IA to the extent that the SH2 domains resemble two domains connected by a flexible linker, the affinity of phosphorylated tSH2 for receptor dp-ITAM is not reduced to monofunctional binding of a single SH2 to one pYXX(I/L) cassette, contrary to previous expectation. A multistate-binding analysis of ITAM association with tSH2 that accounts for discrete binding to each SH2 domain is found an effective approach for defining how IA Y130 phosphorylation alters the equilibrium of receptor binding. Together based on the binding analysis involved multiple complexes, we conclude the dramatic decrease of the apparent binding affinity after Y130 phosphorylation is largely an effect on the intra-molecular binding step after the initial inter-molecular association of phosphorylated tSH2 with dp-ITAM.

Results

Interdomain A phosphorylation partially, but not fully decouples the two SH2 domains

The binding affinity of dp-ITAM to Syk depends on the optimal separation and relative orientation of the tandem SH2 domains observed in the crystallographic structure (Fig. 1) to enable bifunctional binding of two YXX(I/L) cassettes simultaneously across both SH2 domains. To probe how IA phosphorylation alters the conformational ensemble of domain structure, we determine the overall rotational correlation times of the individual SH2 domains from NMR heteronuclear relaxation to characterize the extent of interaction between the SH2 domains in tSH2, a 28 kDa construct of Syk comprising the two SH2 domains plus IA. Strong interactions between (N)SH2 and (C)SH2 would result in a tightly coupled structure, manifest by a longer correlation time consistent with the overall 'rigid-body' structure defined by crystallography, while weakened SH2–SH2 domain coupling would be reflected by a shorter correlation time.[18] The question addressed here is whether the SH2-domain structure of Syk phosphorylated on Y130 is one in which the two SH2 domains are only partly decoupled because phosphorylated IA retains some structural order, as would be necessary to hold the two binding pockets at a distance too long for bifunctional binding. Alternatively, are the SH2 domains fully decoupled due to a randomly disordered IA linker enabling independent rotational motion of the two SH2 domains, as was suggested from an alternative computational analysis[25] of the correlation times that opposed the original interpretation?[18] NMR relaxation rates are compared for three proteins: unphosphorylated tSH2, tSH2$_{PM}$, which is the same construct but with the single amino-acid substitution Y130E to mimic phosphorylation, and tSH2$_{FX}$, which is a construct with IA residues 119–162 replaced by a 20-amino-acid flexible linker, (GGS)$_3$GS(GGS)$_3$. The Y130E substitution of tSH2$_{PM}$ has been used in cells to mimic phosphorylation,[21,26,27] and is reasonable given that Y130 is well solvated with few intra-molecular contacts consistent with the notion that phosphorylation serves to mediate the conformational ensemble of Syk domain structure rather than by a direct structural interaction. The (GGS)$_3$GS(GGS)$_3$ linker of tSH2$_{FX}$ was shown to be flexible and enables independent rotation of connected globular domains.[28]

^{15}N heteronuclear relaxation rates and NOE values measured at 800 MHz field strength for residues from tSH2, tSH2$_{PM}$, and tSH2$_{FX}$ are shown in Fig. 2. Values are not shown for IA residues of tSH2 because these resonances are exchange broadened and unobservable, nor for IA resonances of tSH2$_{FX}$ because of degeneracy. The NOE values from SH2 residues of all proteins are similar and near the long correlation-time limit of 0.83, while the NOEs for IA resonances of tSH2$_{PM}$ are close to 0.5 reflecting greater internal mobility of this linker region. The long correlation-time limit occurs near 4 ns at 800 MHz and thus the SH2 domains are folded and well-ordered for the three proteins. In contrast, distinct residue-profiles for T_1, T_2 and T_1/T_2 values are observed for each of the three proteins. The relaxation-rate values for tSH2$_{PM}$ are intermediate to those for tSH2 and tSH2$_{FX}$, and thus the hydrodynamic rotation of the SH2 domains of tSH2$_{PM}$ is intermediate and not at the limit of tSH2$_{FX}$ with a flexible linker between the two SH2 domains.

To quantify the hydrodynamic differences, rotational diffusion correlation times, τ_c, were obtained for tSH2, tSH2$_{PM}$ and tSH2$_{FX}$ by fitting the relaxation data for residues that have minimal internal motion and from an individual SH2 domain using the program TENSOR 2.0[29] (see Methods). The T_1/T_2 values and estimated τ_c are summarized in Table 1. There is a good agreement between the T_1/T_2 and τ_c values for (N)SH2 and (C)SH2 resonances. Similar to previous results,[18] τ_c estimated from experimental relaxation data of tSH2 is equal to 17 ± 1 ns for both SH2 domains, which is the same value predicted from the X-ray structure (17 ± 1 ns for the six molecules in the asymmetric unit of PDB ID 1A81) using the program HYRDONMR.[30] The dp-ITAM bound crystal structure of tSH2 is used to interpret the relaxation data rather than an unbound crystal structure in accord with Zhang, et al.[18] as noted above. Similar τ_c shows that the SH2 domain–domain interaction is strong and the two domains are tightly coupled and rotate in solution as a rigid unit. For the other limit, tSH2$_{FX}$ with a flexible linker has a much smaller τ_c equal to 8.2 ± 0.4 and 8.7 ± 0.4 ns determined from (N)SH2 and (C)SH2, respectively. These values are similar to 8.1 ± 0.4 ns obtained for N-SH2 + IA, a construct comprising a single SH2 domain plus the linker residues for IA, so that the SH2 domains of tSH2$_{FX}$ are found to rotate independently as expected for a flexible linker. By contrast, the τ_c value determined for tSH2$_{PM}$ is 11 ± 1 ns for both SH2 domains, intermediate to tSH2 and tSH2$_{FX}$, suggesting that the linker of tSH2$_{PM}$ is not as structurally ordered as in tSH2, but not fully disordered as in tSH2$_{FX}$.

In summary, phosphorylation increases conformational disorder of linker A causing the SH2 domains to rotate in solution like a partially coupled unit, as suggested previously by Zhang et al.[18] but contrary to an alternative interpretation by others[25] that the correlation times indicate the SH2 domains rotate independently as the beads-on-a-string model of polymers. Therefore the two SH2 domains of Syk phosphorylated in IA could be conformationally restricted such that the two pY binding pockets could not support bifunctional binding.
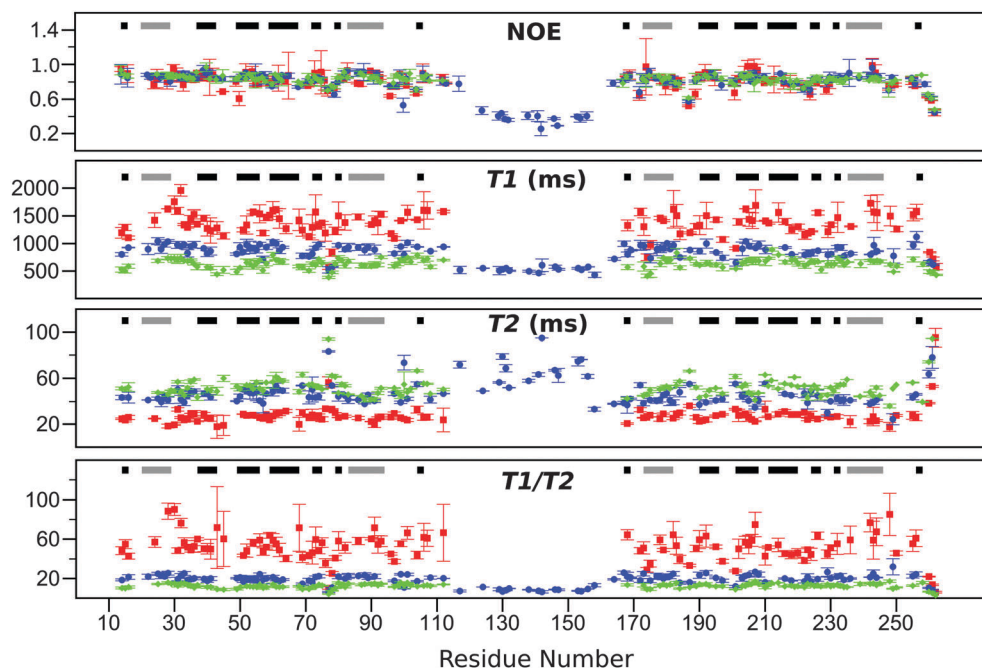
This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 5807–5818 | 5809

Fig. 2 Comparison of NOE, T_1, T_2 and T_1/T_2 values for Syk tSH2 constructs. Red: tSH2; blue: tSH2$_{PM}$; green: tSH2$_{FX}$. Data points represent the mean values and the error bars indicate standard deviations from two independent experiments. The bars on top denote secondary structure elements. Grey: α-helices. Black: β-sheets. The secondary structure elements are as follows: βA (residues 14–16, 167–169), βB (residues 37–43, 190–196), βC (residues 49–56, 201–208), βDβDʹ (residues 59–68, 211–220), βE (residues 72–75, 224–227), βF (residues 79–81, 231–233), βG (residues 104–106, 256–258), αA (residues 20–29, 173–182), αB (residues 83–94, 235–246). The resonances of IA are not observed in tSH2 and are degenerate in tSH2$_{FX}$.

Table 1 Comparison of T_1/T_2, and correlation time (τ_c) for different protein constructs

Construct	Domain[c]	T_1/T_2[d]	τ_c[e] (ns)	τ_{pred}[f] (ns)
tSH2	(N)SH2	53 ± 7	17 ± 1	17 ± 1[g]
tSH2	(C)SH2	52 ± 10	17 ± 1	17 ± 1[g]
tSH2$_{PM}$	(N)SH2	21 ± 2	11 ± 1	—
tSH2$_{PM}$	(C)SH2	22 ± 3	11 ± 1	—
tSH2$_{FX}$	(N)SH2	12 ± 2	8.2 ± 0.4	—
tSH2$_{FX}$	(C)SH2	13 ± 2	8.7 ± 0.4	—
N-SH2 + IA[a]	—	13 ± 1	8.1 ± 0.4	—
N-SH2[b]	—	6.4 ± 1.1	5.2 ± 0.3	5.9[h]

[a] Isolated N-terminal SH2 domain plus interdomain A, residues 8 to 162. [b] Isolated N-terminal SH2 domain. [c] Resonances from the individual domain in tSH2 used for averaging T_1/T_2 and calculation of τ_c. [d] Data are shown as mean ± SD of the qualified residues (>20 residues) from each domain; values of the individual residues are means from two independent measurements. Only residues with low internal mobility and free from conformational exchange are used for this summary and τ_c calculation; see Methods for definition. [e] Correlation time computed using TENSOR 2.0 from experimental NMR measurements for each domain; data are shown as mean ± SD of the fitted values from two independent experiments. [f] Computationally predicted correlation time using HYDRONMR with the crystal structure (PDB ID 1A81, radius of the atomic elements α = 1.5 Å, 298 K). [g] Mean ± SD of the predicted correlation time for the six molecules in the asymmetric unit of the crystal structure. [h] The predicted correlation time for the isolated C-terminal SH2 domain is 5.4 ns, similar to that of the isolated N-terminal SH2 domain.

Characterizing the interaction of Syk tandem SH2 with ITAM by NMR

Given that the SH2 domains in tSH2$_{PM}$ are not rotationally independent, it is reasonable that IA phosphorylation alters the

domain conformational ensemble such that the SH2 domain separation is restricted to a distance too large to allow an ITAM molecule to associate bifunctionally. The reduced affinity of IA-phosphorylated Syk would therefore approximate the monofunctional affinity of a single SH2 domain. We therefore investigated the binding of Syk tandem SH2 proteins to peptides with one pYXX(I/L) cassette in comparison to a dp-ITAM peptide with two phosphotyrosines. The measurements reported here are for a dp-ITAM peptide derived from the CD3ε chain of the T cell receptor, ITP, which has a reported affinity for Syk tSH2 of 20–30 nM.[11,13] The two half peptides of ITP, each containing only one pY, are from the N-terminus, N-IHP, and the C-terminus, C-IHP. Peptide sequences are given in Fig. 1c.

The titration of tSH2 proteins with ITP, N-IHP and C-IHP was monitored by changes in $^1H^{15}N$ HSQC chemical shifts and line shapes. An advantage for investigating molecular interactions using NMR rather than other biophysical methods, such as gel filtration, equilibrium dialysis, calorimetry and fluorescence, is the residue-level information obtained compared to a single readout to follow binding. Of particular importance for the tSH2 system is that binding to a specific SH2 domain, (N)SH2 or (C)SH2, can be followed independently. Several resonances in each SH2 domain are well resolved in both the free and bound states and were used to monitor binding: G32, L37, L52, H61, Y73, A74, I99 and F106 for (N)SH2, and D175, G184, L192, C205, G210 and S244 for (C)SH2. The titration profiles within one SH2 domain were equivalent within experimental error and could be considered independent measurements of binding to that domain (see ESI†).
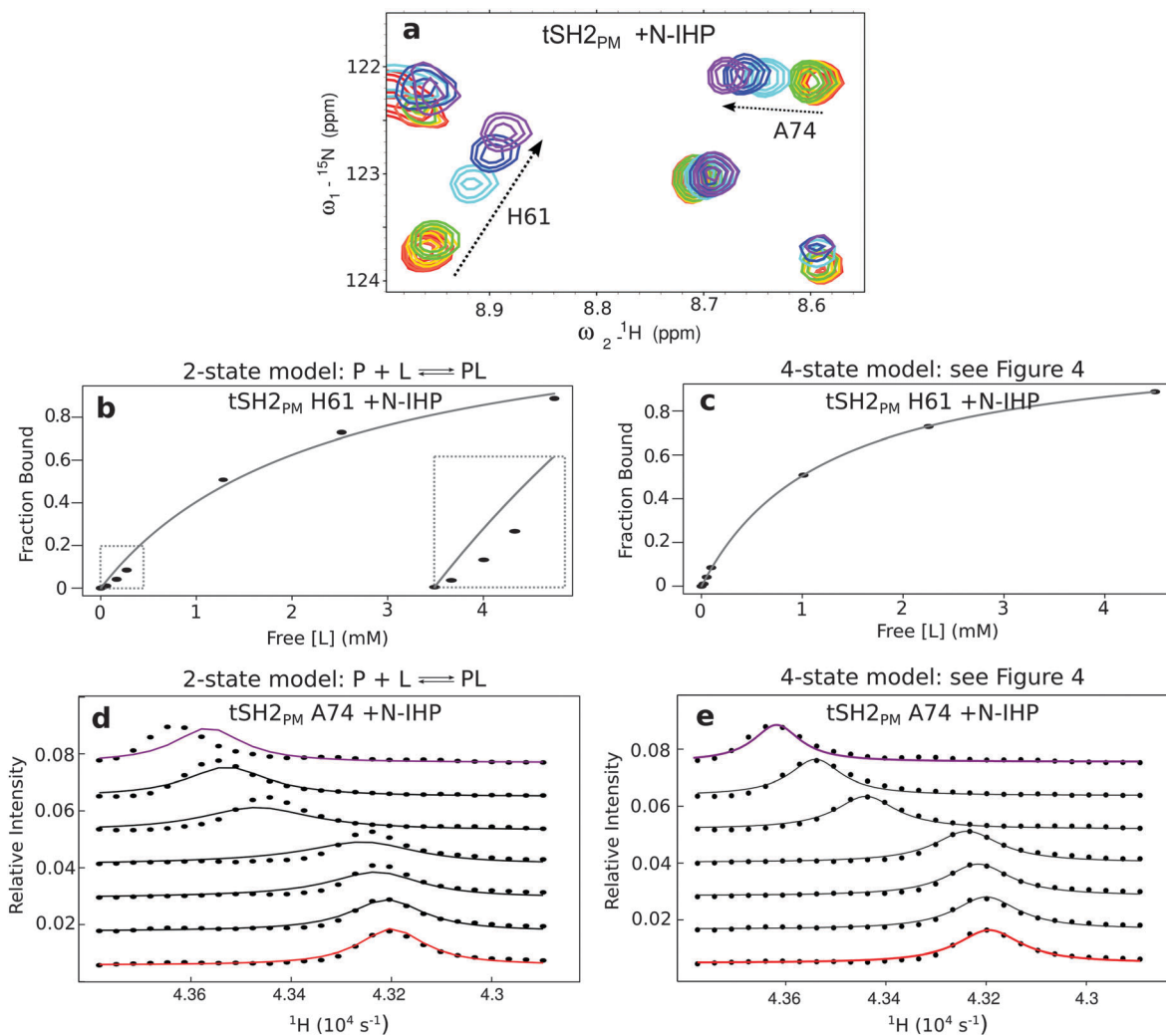
Fig. 3 Representative NMR ^{15}N-HSQC chemical shift titration data and comparison of fitting the 2-state model and 4-state model by titration-curve/line-shape analysis for IHP complexes. (a) Overlaid HSQC spectra for tSH2$_{PM}$ (0.3 mM) titration with N-IHP. Increasing [N-IHP] corresponds to color changes from red, orange, yellow, green, cyan, blue, to purple. A zoomed region with H61 and A74 is shown as an example. The arrows indicate the shift of peaks during the titration process. (b and c) Titration-curve analysis for a fast-exchange residue, H61, in tSH2$_{PM}$ plus N-IHP with either the conventional 2-state model or the 4-state model in Fig. 4. The fraction bound values are the peak centers after normalization, and the free ligand concentrations are calculated with the indicated model. Dotted box: zoomed view. Circles: raw data. Line: predicted values with the indicated model. (d and e) Line-shape analysis for a fast exchange residue, A74, in tSH2$_{PM}$ plus N-IHP with either the conventional 2-state model or the 4-state model. Circles: raw line-shape data. Lines: predicted line-shape with the indicated model (red: the first titration point with no ligand. Purple: the last titration point with high concentration of ligand). The 4-state model fits well with the NMR data while the 2-state model does not.

NMR titration data were analyzed in the form of either binding curves or spectral line shapes,[31–34] as described in detail in ESI.† Binding curves were generated from changes in the peak position for resonances in fast or fast-intermediate exchange on the NMR chemical-shift time scale based on obtaining hyperbolic binding curves as a function of free ligand concentration using fitted values of the equilibrium binding constants. In this exchange regime, the peak position approximates well the fraction of bound species, and the exchange rate is greater than approximately 2× the frequency difference between the free and bound states of the protein.[35] In contrast, the analysis of NMR line shapes utilizes the whole peak envelope and is not limited to resonances in the fast or fast-intermediate exchange regime. Line shape analysis was applied

to all monitored resonances, including those in the slow-intermediate or slow regime for cases of relatively higher binding affinity. Binding rate constants, as well as equilibrium binding constants were estimated from line-shape analysis.

Multi-site analysis for interactions of Syk tSH2 proteins with half ITAM peptides

The interactions between Syk tSH2 proteins (tSH2, tSH2$_{PM}$, or tSH2$_{FX}$) and the half ITAM peptides (N-IHP or C-IHP) were investigated to determine the single SH2-domain binding affinity. These monofunctional interactions were found to be low affinity, corresponding to the fast or fast-intermediate exchange regime, and thus both titration-curve and line-shape methods were used to extract the binding constants.

The NMR titration-curves and line-shapes were initially fit assuming a simple 2-state binding model: binding curves were fit using R[36] (home-written scripts utilizing DRC package[37]) and line shapes were fit using LineShapeKin[35] (version 3.2, http://lineshapekin.net). Nevertheless, the best-fit 2-state model cannot explain the experimental data. The poor fit is illustrated with data from resonances of H61 and A74 in Fig. 3b and d, respectively. Data for H61 plotted as fraction bound vs. free [L] using the best-fit value for K_D show sigmoidal behavior and deviate systematically from the theoretical curve (Fig. 3b). Data for A74 illustrates the poor agreement between the experimental line shape and the predicted line shape (Fig. 3d). The poor fit of the NMR data indicate a model with more states is needed, which is reasonable given the protein has two potential binding sites, one on each SH2 domain, and the head-to-tail binding mode of Syk (N)SH2 and (C)SH2 binding to the C-terminal and

N-terminal pYXX(I/L) cassette, respectively, is not absolute for the half ITAM peptides.

A 4-state model accounts for the data on the interactions between the tSH2 proteins and an IHP ligand. As shown in Fig. 4, the model contains four states for a tSH2 protein interacting with an IHP ligand: the unbound state, NC where N and C represent (N)SH2 and (C)SH2, respectively, the two states with one ligand bound to either SH2 domain, N^LC and NC^L, and the 2:1 complex with two ligands bound, N^LC^L. An IHP can bind to an SH2 domain with the equilibrium dissociation constant,

$$K_N = \frac{[NC][L]}{[N^LC]} = \frac{[NC^L][L]}{[N^LC^L]}$$
$$K_C = \frac{[NC][L]}{[NC^L]} = \frac{[N^LC][L]}{[N^LC^L]}$$
(1)

where L denotes an IHP ligand (N-IHP or C-IHP) with only one pYXX(I/L) cassette. To determine K_N and K_C, independent measurements for the IHP peptide association with (N)SH2 and (C)SH2 domains are needed and was possible with NMR by global fitting of peaks from the separate domains (see ESI† for details). The improved fitting of the binding curves and line shapes obtained using the 4-state model is illustrated for resonances of H61 and A74 by comparing Fig. 3c with b and e with d.

Fitting results for the thermodynamic parameters of binding the IHPs to the three tSH2 proteins are summarized in Table 2. The dissociation constants K_N and K_C determined from fitting either the binding curves or the line shapes are in reasonable agreement. Values from the two methods are within a factor of 1.5 of each other, and do not vary systematically. It is observed that N-IHP or C-IHP bind a given SH2 domain in the three tSH2 proteins with similarly equivalent dissociation constants. Thus, the linker state does not affect the intrinsic binding affinity of a single SH2 domain with the half ITAM, as expected.

The single SH2 domain affinities listed in Table 2 are unexpectedly low for the binding of a phosphotyrosine peptide to an SH2 domain. Other SH2 peptide complexes typically have dissociation constant values ranging from 20 nM to 50 μM (ΔG^0 from −5.9 to −10.6 kcal mol^{-1}, SH2 domains from nine proteins).[38–40] The affinity previously reported from an analytical ultracentrifugation method for Syk tSH2$_{PM}$ binding dp-ITAM was 1.8 μM,[18] and therefore the notion that IA phosphorylation
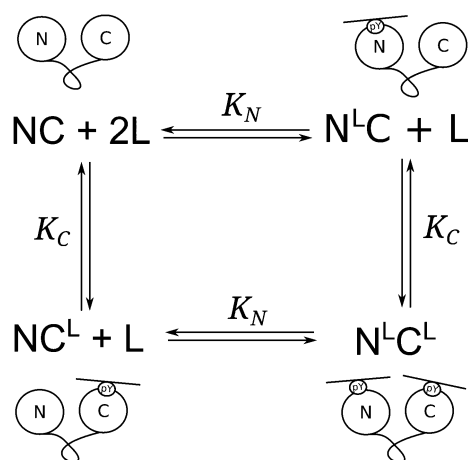

Fig. 4 The 4-state model for the interactions between a tSH2 protein and an IHP ligand with one pYXX(I/L) cassette. The possible tSH2 molecules are tSH2, tSH2$_{PM}$, or tSH2$_{FX}$. N and C represent (N)SH2 and (C)SH2, respectively. L denotes an IHP ligand with a single pYXX(I/L) cassette; the possible IHP ligands are N-IHP or C-IHP (see Fig. 1). The four states for the tSH2 molecule are the ligand-free state (NC), the two intermediate states with only one ligand bound to either the (N)SH2 domain (N^LC) or the (C)SH2 domain (NC^L), and the 2:1 complex which has a ligand bound to each of the two domains (N^LC^L). Assuming the two domains bind independently, then the system has only two dissociation constants, K_N and K_C for each domain, respectively.

Table 2 Dissociation and kinetic constants determined for individual SH2 domains of different IHP complexes with the 4-state model. The values for the SH2 + IHP interaction relevant to the head-to-tail binding of dp-ITAM to tSH2 are shown in bold

Complex	(N)SH2				(C)SH2			
	$K_N{}^a$ (μM)	$K_N{}^b$ (μM)	$k_{on}^N{}^b$ (10^6 s^{-1} M^{-1})	$k_{off}^N{}^b$ (s^{-1})	$K_C{}^a$ (μM)	$K_C{}^b$ (μM)	$k_{on}^C{}^b$ (10^6 s^{-1} M^{-1})	$k_{off}^C{}^b$ (s^{-1})
tSH2/N-IHP	450 ± 30	730 ± 20	2.5 ± 0.5	1800 ± 300	**59 ± 6**	**69 ± 4**	**5.5 ± 0.4**	**380 ± 20**
tSH2$_{PM}$/N-IHP	820 ± 60	1300 ± 100	2.9 ± 0.3	3700 ± 300	**69 ± 3**	**84 ± 4**	**6.6 ± 0.4**	**550 ± 20**
tSH2$_{FX}$/N-IHP	570 ± 20	820 ± 50	1.6 ± 0.3	1300 ± 200	**50 ± 3**	**53 ± 7**	**6.6 ± 1.0**	**350 ± 20**
tSH2/C-IHP	**510 ± 150**	**360 ± 40**	**2.6 ± 0.3**	**950 ± 60**	350 ± 60	220 ± 40	12 ± 5	2600 ± 900
tSH2$_{PM}$/C-IHP	**540 ± 30**	**610 ± 110**	**3.3 ± 1.4**	**2000 ± 800**	280 ± 20	310 ± 50	13 ± 11	4100 ± 3300
tSH2$_{FX}$/C-IHP	**570 ± 200**	**510 ± 90**	**3.9 ± 1.3**	**2000 ± 600**	350 ± 130	290 ± 50	6.2 ± 3.3	1800 ± 900

^{a} Dissociation constants determined by titration-curve analysis. ^{b} Dissociation and kinetic constants determined by line-shape analysis. Data are shown as mean ± SD of fitted values from two independent experiments.

5812 | *Phys. Chem. Chem. Phys.*, 2016, **18**, 5807–5818

This journal is © the Owner Societies 2016

inhibits the Syk–receptor ITAM interaction by reducing the association to a single SH2 binding event is challenged by the significantly lower affinity of monofunctional SH2 binding determined here (Table 2).

Examination of Table 2 finds that (C)SH2 associates with N-IHP with an order of magnitude smaller dissociation constant ($K_C \sim$ 50–70 μM) than that of the other three half-ITAM complexes ($K_x \sim$ 300–800 μM). That (C)SH2 associates with N-IHP with higher affinity is of biological interest as it relates to the observation that a Src-family kinase phosphorylates the N-terminal tyrosine of ITAM only[1,10] while Syk is thought to phosphorylate the C-terminal tyrosine.[1] Given the relative dissociation constants reported here, it follows that Src phosphorylation of the N-terminal tyrosine provides a docking site for Syk C(SH2), and so facilitates Syk in the subsequent phosphorylation of the C-terminal tyrosine of ITAM to generate the doubly phosphorylated ITAM for a stronger membrane anchoring of Syk to sustain membrane signaling.

With the fitted binding constants, the continuous change of concentration of each species in the 4-state model was simulated for each tSH2 construct binding with each IHP (see ESI†). The three tSH2 constructs behaved similarly when binding to the same IHP because the single-domain binding affinities are similar as aforementioned. As expected, most tSH2 molecules form the two intermediate states (with only 1 IHP ligand bound; N^LC and NC^L in Fig. 4) at the beginning of titration (the NC^L state has higher population due to its higher affinity, especially when binding with N-IHP); however, the concentration of these two intermediate states decreases as the concentration of ligand increases; Finally, the population of the 2:1 complex (N^LC^L in Fig. 4) dominates at very high ligand concentrations.

Stepwise association of tSH2 proteins with full-length dp-ITAM

The domain-specific information provided by NMR enabled analysis of the effect the domain-structure ensemble and phosphorylation of IA on the energetics of dp-ITAM binding across two SH2 domains. NMR lineshape analysis was used to measure the association of the three Syk tandem SH2 proteins with the full-length ITP comprising two pYXX(I/L) cassettes. The IHP binding free energies from the results given in Table 2 indicate a complete binding scheme should account for the interaction of both ITP pYXX(I/L) cassettes with either SH2 domain and thus the simplest yet complete model includes ten states (Fig. 5a). The first binding step of the free tSH2 protein, NC, with ITP leads to one of four 1:1, monofunctional complexes with a single SH2 domain bound to one pYXX(I/L) cassette of ITP: N^NC, N^CC, NC^N and NC^C. (Superscript N and C refer to ligation by the N-terminal and C-terminal cassette of ITP, respectively.) For N^CC and NC^N with head-to-tail association, the binding event for the second SH2 domain is either intra-molecular (isomerization) to form the bifunctional complex $N'C'$, or inter-molecular to form a 2:1 ITP:tSH2 ternary complex. If the first binding event is not a head-to-tail association, then the intra-molecular event is assumed improbable and the second domain is occupied only upon binding a second ITP molecule. The four 2:1 ternary complexes after a second inter-molecular binding step are N^NC^N, N^CC^N, N^NC^C, and N^CC^C.
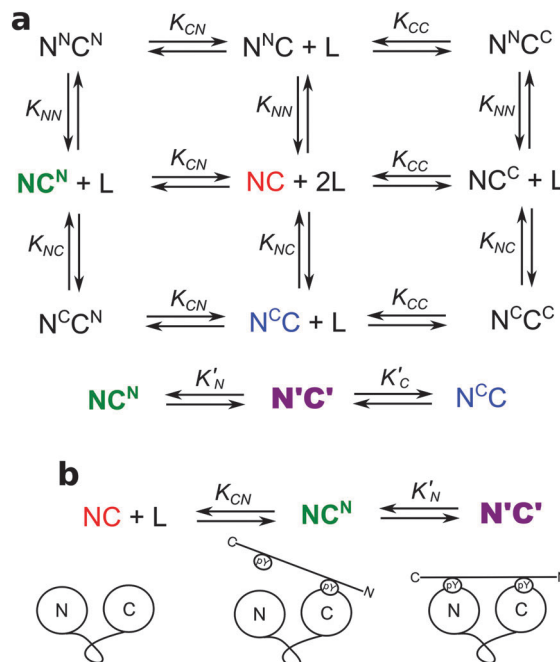

Fig. 5 The 10-state and 3-state models for the interactions between a tSH2 molecule and the full-length ITP comprising two pYXX(I/L) cassettes (see Fig. 1). The possible tSH2 molecules are tSH2, tSH2$_{PM}$, or tSH2$_{FX}$. L denotes the full-length ITP with two pYXX(I/L) cassettes. (a) The 10-state model with all possible interactions between an SH2 domain (N or C) of tSH2 and the N- or C-terminal cassette (superscript N or C) of the full-length ITP. The first step is inter-molecular binding: a tSH2 molecule in the ligand-free state (NC, shown in red) binds either the N- or C-terminal cassette of ITP through either of its two domains to form a 1:1 mono-functional complex (N^NC, N^CC, NC^N, or NC^C). The second step can be another inter-molecular binding; the 1:1 monofunctional complex can bind to another ITP through its unbound domain to form a 2:1 ternary complex (N^NC^N, N^CC^N, N^NC^C, or N^CC^C). Alternatively, the second step can be an intra-molecular binding (isomerization) whereby N^CC or NC^N (shown in blue and green, respectively) form the 1:1 bifunctional complex $N'C'$ (shown in purple; this is the head-to-tail binding complex; see Fig. 1). Assuming each SH2 domain binds an ITP peptide independently, the model has four equilibrium dissociation constants for the inter-molecular binding events: K_{NN}, K_{NC}, K_{CN}, and K_{CC}, where the first and second subscript indicates the SH2 domain and the ITP cassette, respectively. The model also has two equilibrium isomerization constants for the intra-molecular bindings: K'_N for NC^N isomerization to $N'C'$, and K'_C for N^CC isomerization to $N'C'$. NC^N and $N'C'$ (shown in bold) are predicted to be the major products of the first and second binding steps, respectively. (b) The simplified 3-state model. Considering only the major products of each step, the 10-state model can be simplified to this 3-state model. The structure of each state is shown below the name.

Each inter-molecular binding of an ITP pYXX(I/L) cassette with an SH2 domain is assumed to be independent. The equilibrium of (C)SH2 binding the N-terminal cassette of ITP is defined with the equilibrium dissociation constant K_{CN},

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

where L denotes an ITP ligand which has two pYXX(I/L) cassettes. Analogous expressions define K_{NN}, K_{NC}, and K_{CC}. The equilibrium between tSH2 with ITP bound monofunctionally to a single SH2

domain and bound bifunctionally to both SH2 domains is described by the concentration-independent equilibrium isomerization constants, K'_N and K'_C,

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

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

(3)

Based on the affinities determined for an SH2 domain to bind either N-IHP or C-IHP with a single pYXX(I/L) cassette (Table 2), the major species is NC^N after the first binding step, and $N'C'$ after the second binding step at sub-mM concentrations. Accordingly, it is

worthwhile to consider a simpler 3-state model (Fig. 5b) along with the 10-state model. The first step of the reduced 3-state model is the (C)SH2 domain binding inter-molecularly with the N-terminal pYXX(I/L) cassette of ITP. The second step is the intra-molecular association (isomerization) of the (N)SH2 domain with the C-terminal pYXX(I/L) cassette of the same ITP molecule.

The NMR spectra for the titration of tSH2, tSH2$_{PM}$ and tSH2$_{FX}$ with ITP showed slow or intermediate exchange behavior as a result of higher binding affinities than N-IHP and C-IHP. An example of the slow-intermediate exchange behavior of resonances with larger chemical shift perturbation (CSP) and dual peak centers in the intermediate titration points is given in Fig. 6a. For peaks showing smaller CSP, a single peak was observed but the peak
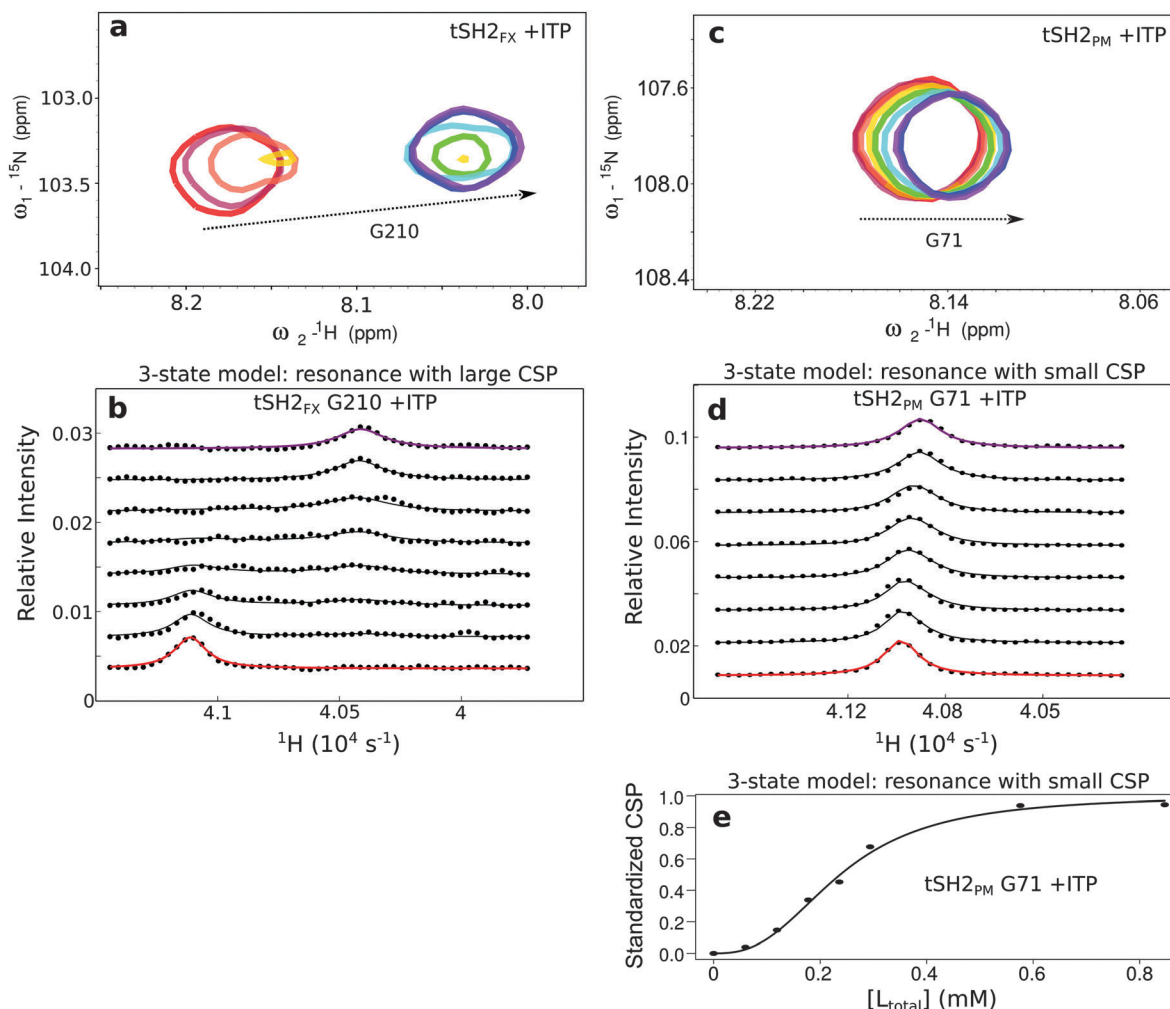

Fig. 6 Representative NMR titration data and the line-shape analysis for ITP complexes showing the slower-exchange behavior than that of IHP complexes. (a) Overlaid HSQC spectra for tSH2$_{FX}$ (0.3 mM) titration with ITP. Increasing [ITP] corresponds to color changes from red, orange, yellow, green, cyan, blue, to purple. A zoomed region with G210, which has relatively large CSP, is shown as an example. The arrow indicates the direction of change during the titration process. (b) Line-shape analysis for a slow-intermediate exchange residue, G210, in tSH2$_{FX}$ plus ITP. Dotted lines: raw line-shape data. Solid lines: predicted line-shape with the 3-state model (red: the first titration point with no ligand. Purple: the last titration point with high concentration of ligand). (c) Similar overlaid HSQC spectra for tSH2$_{PM}$ (0.3 mM) titration with ITP showing G71, which has small CSP. (d) Line-shape analysis for G71 in tSH2$_{PM}$ plus ITP with the 3-state model. Note that there is no significant line broadening due to the small CSP; however the apparent exchange rates are comparable to the difference in frequency so that exchange is in the intermediate regime. (e) Plot of the standardized CSP vs. [L$_{total}$] for the intermediate-exchange residue, G71, in tSH2$_{PM}$ plus ITP. Circles: raw data. Line: predicted values with the 3-state model. Note the sigmoidal behavior due to intermediate exchange. Many residues in the ITP complexes showed this intermediate or slow-intermediate exchange behavior which prevents the titration-curve analysis because the shifts of peak centers no longer represent the fractions bound.

5814 | *Phys. Chem. Chem. Phys.*, 2016, **18**, 5807–5818

This journal is © the Owner Societies 2016

Table 3 Intra-molecular binding constants (see Fig. 5 for definition) and apparent equilibrium dissociation constants determined for ITP complexes of tSH2, tSH2$_{PM}$ and tSH2$_{FX}$

Complex	3-state model		10-state model		
	K_N' (10^{-3})	$K^{App\,a}$ (μM)	K_N' (10^{-3})	$K_C'^{\,b}$ (10^{-3})	$K^{App\,c}$ (μM)
tSH2/ITP	$\ll 5$	$\ll 0.3$	$\ll 4$	$\ll 0.7$	$\ll 0.3$
tSH2$_{PM}$/ITP	38 ± 8	3.2 ± 0.7	32 ± 10	4.4 ± 1.6	2.7 ± 0.8
tSH2$_{FX}$/ITP	120 ± 20	6.1 ± 1.3	100 ± 10	11 ± 3	5.5 ± 1.1

^{a} $K^{App} = K_{CN} \times K_N'$. ^{b} $K_C' = K_{CN} \times K_N'/K_{NC}$. ^{c} $K^{App} = K_{CN} \times K_N' = K_{NC} \times K_C'$. See Table 2 for the values of K_{CN} and K_{NC} determined from IHP complexes (line-shape analysis). Data are shown as mean $\pm$ SD or detection limit of fitted values from three independent experiments.

centers no longer represent the fraction bound (Fig. 6c); the position of the peak center plotted against ligand concentration displays a strong sigmoidal character due to intermediate exchange (Fig. 6e). Both types of residues were analyzed with the line-shape method (Fig. 6b and d).

The line-shape analysis for the tSH2/ITP complexes was performed using both the 10-state and 3-state models (Fig. 5). A fit for all thermodynamic parameters is not possible, and thus the equilibrium dissociation constants of the inter-molecular binding steps were fixed to the values obtained from the line-shape analysis of the IHP complexes (Table 2), assuming the binding free energy of a single SH2 domain to the full-length ITP is approximately equal to that of IHPs. Fittings with either the 10-state or 3-state models gave equally good agreement with the experimental line shapes (Fig. 6b), and similar values for the equilibrium constants of the intra-molecular binding steps, K_N' and K_C' (Table 3). The 10-state model defines one more equilibrium parameter than the 3-state model: K_C', the equilibrium dissociation constant for the intra-molecular binding between (C)SH2 domain and N-terminal pYXX(I/L) cassette. The off-rate values of the intra-molecular binding steps did not converge upon fitting and cannot be determined for the three ITP complexes due to lack of direct observable signal of the intermediate states. As might be expected, the intermediate states for tSH2 constructs binding with dp-ITAM do not have chemical shifts distinct from the final bound species based on the fact that all of the resonances showed linear titration patterns.[41] This inability to directly observe the intermediate state prevents determination of off-rate values.

Most relevant to the regulation of the Syk–receptor interaction is the estimate for K_N', the equilibrium dissociation constant for the intra-molecular binding step of (N)SH2 to the C-terminal pYXX(I/L) cassette. The fitting of the 3-state model (Fig. 5b) yields K_N' values of $\ll 0.005$, $(38 \pm 8) \times 10^{-3}$, and $(120 \pm 20) \times 10^{-3}$ for tSH2, tSH2$_{PM}$, and tSH2$_{FX}$, respectively. The intra-molecular equilibrium (isomerization) favors bifunctional binding for all three forms of tSH2, even in the case of tSH2$_{FX}$ with the flexible linker sequence between the two SH2 domains. For unphosphorylated tSH2 with the strongest SH2–SH2 interaction, the second intra-molecular binding step is too favorable to define K_N' accurately using line-shape analysis. Bifunctional binding is favored somewhat more

strongly for tSH2$_{PM}$ than tSH2$_{FX}$ (2 to 3-fold stronger, $p < 0.05$ based on two-sided t-test).

The macroscopic apparent binding affinity based on the 3-state model ($K^{App} = K_{CN} \times K_N'$) is $\ll 0.3$, 3.2 ± 0.7, and 6.1 ± 1.3 μM for tSH2, tSH2$_{PM}$, and tSH2$_{FX}$, respectively. The apparent equilibrium dissociation constant of tSH2 and ITP is expected to be substantially lower than the limits achieved here. Reported values for the apparent K_d are ~ 0.001 to 30 nM from other methods,[11–16] which is higher affinity than can be determined by direct NMR measurements.[41]

The concentration of each species over the course of the titration was simulated using fitted thermodynamic constants for the three tSH2/ITP complexes (see ESI†). For the 3-state model, concentrations of both the intermediate 1:1 monofunctional complex (NC^N) and the final 1:1 bifunctional complex (N'C') increase as the ligand concentration increases, and ratio of [NC^N]:[N'C'] remains the same (equals to K_N') due to the first-order reaction nature of the intra-molecular binding step (isomerization). tSH2$_{FX}$/ITP showed higher concentration of the intermediate state than the other two due to its higher value of K_N'. The 3-state model of course is not accurate at high peptide concentrations. In comparison, the 10-state model behaved similarly as the 3-state model at ligand concentrations less than ~ 450 μM (total protein concentration is 300 μM); however, as the ligand concentration increases above ~ 900 μM, the formation of the 2:1 complex is favored, particularly for tSH2$_{FX}$/ITP due to the higher value of K_N'. These results of the 10-state model are more realistic than the 3-state model at high ligand concentrations, and may be required under certain physiological conditions as Syk was reported to form 2:1 complexes with the single-YxxL-containing receptors, CLEC-2.[42]

Conclusion

The allosteric regulation of Syk binding to dp-ITAM regions of membrane receptors derives from the altered SH2–SH2 domain structure upon phosphorylation of the IA linker connecting the two SH2 domains. From a comparison of heteronuclear relaxation properties of tSH2, tSH2$_{PM}$, and tSH2$_{FX}$, the phosphorylation of IA at Y130 is found to weaken the SH2–SH2 interaction by inducing disorder in IA. Nonetheless, the SH2 domains are not rotationally independent to the extent of being joined by a fully flexible linker. The change in SH2–SH2 domain structure indicated by the reduction in τ_c cannot be accounted for by the differences among the six molecules in the asymmetric unit of dp-ITAM bound tSH2 (PDB ID 1A81) or the alternative SH2–SH2 domain structure in unbound full-length Syk (PDB ID 4FL2).

Site-specific free energies of binding determined from the residue-level information of NMR spectroscopy enable a description of the energetics for the separate binding to (N)SH2 and (C)SH2 domains that gives insight into the allosteric regulation of Syk by phosphorylation of IA. The change in the SH2–SH2 conformational ensemble upon introduction of negative charge at position Y130 alters the binding process but does not restrict association to a single SH2 domain. Bifunctional binding was

This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 5807–5818 | **5815**

established for the association of the dp-ITAM peptide ITP with tSH2$_{PM}$ negatively charged at position Y130 in IA, as well as tSH2$_{FX}$ with a fully flexible linker. Given the flexibility of IA in tSH2$_{PM}$ and tSH2$_{FX}$, which abolishes the stable domain structure matching the pYXX(I/L) spacing of ITAM, it was not apparent from the outset that tSH2$_{PM}$ or tSH2$_{FX}$ would engage in bifunctional binding in preference to association with two ITP ligands. The energetics of the site-specific association of each pYXX(I/L) cassette to separate SH2 domains (Table 2) establishes that for all three tSH2 proteins the binding free energy for bifunctional association of dp-ITAM is favored over forming the ternary complex with two inter-molecular binding steps (Table 3) at typical ligand concentrations (ligand : protein ratios less than 20 : 1; see ESI†). The reduced affinity of Syk phosphorylated at Y130 for receptor ITAM is accounted for by a less favorable isomerization equilibrium to form the bifunctional complex N′C′. As a result, an important finding of this work is that IA phosphorylation alters the partitioning of the second association process between the bifunctionally associated ITAM–Syk complex and the ternary complex formed from two inter-molecular binding events. For tSH2$_{PM}$, the population of the ternary complex increases at concentrations that might approach those of clustered receptors (*i.e.* above 20 : 1). The low intrinsic affinity of a single Syk SH2 domain for phosphotyrosine-containing peptides is a key factor in the ability to regulate this partitioning; the affinity of a single SH2 domain binding must match the isomerization equilibrium to allow a switch from a lower-affinity bifunctional binding to a ternary complex populated at physiologically achievable receptor concentration. Regulation of this partitioning in a feasible concentration regime of clustered receptors would be important for controlling the competition of various membrane receptors or downstream signaling proteins that might associate with Syk through ITAMs with varying separation between pYXX(I/L), or importantly, receptors that have only a half ITAM with a single pYXX(I/L).[42]

Methods

Plasmid construction

cDNA encoding murine Syk tandem SH2 domains, Ser 8 to Gln 264, was cloned into a pET-30a(+) (Novagen) vector[18] for the construct tSH2. The residue Y130 was mutated to glutamate using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) for the construct tSH2$_{PM}$.[18] The construct tSH2$_{FX}$ has interdomain A (Phe 119 to His 162) substituted with a flexible 20-amino-acid linker [(GGS)$_3$GS(GGS)$_3$]. Two PCR reactions were performed for the tSH2$_{FX}$ construct: the cDNA product of reaction 1 encodes (N)SH2 (Ser 8 to Pro 118) plus half of the linker (GGSGGSGGSG), and has cleavage sites for *NdeI* and *BamHI*; the cDNA product of reaction 2 encodes the other half of the linker (SGGSGGSGGS) plus (C)SH2 (Glu 163 to Gln 264), and has cleavage sites for *BamHI* and *XhoI*. These two pieces of cDNA were then cloned into the pET-30a(+) vector: the two inserts were cleaved by *NdeI*, *BamHI*, *and XhoI*; the vector was cleaved by *NdeI*, *EcoRV-HF*, and *XhoI*; ligation was performed using T4 DNA ligase. cDNAs for N-SH2 + IA (Ser 8 to His 162)

and N-SH2 (Ser 8 to Gly 117), with an N-terminal (His)$_6$ tag and TEV cleavage site, were cloned into pET-30a(+) vectors.

Protein expression and purification

Proteins were expressed in *Escherichia coli* strain Rosetta 2 (DE3) (Novagen). Cells were grown at 37 °C in M9 minimal medium containing ^{15}N isotopes to an OD$_{600}$ reading of 1.0 to 1.3. Protein expression was induced by addition of 1 mM of IPTG and the cells were then kept at 18 °C for 18–20 h. Cells were collected by centrifugation and lysed using a French press. The lysate was subjected to ultracentrifugation and the soluble fraction was collected. The proteins of tSH2 constructs were loaded onto a phosphotyrosine-agarose affinity column and eluted with a linear gradient (buffer A: 50 mM Tris, pH 8.0; buffer B: buffer A plus 1 M NaCl). The proteins of N-SH2 + IA and N-SH2 were purified with a HisTrap HP affinity column (GE Healthcare), cleaved with AcTEV protease (Invitrogen), and further purified with the HisTrap HP affinity column.

NMR sample preparation

The purified protein was concentrated to 0.8–1.0 mM with an Amicon Ultra-15 Centrifugal Filter Unit (Millipore) and stored in buffer [50 mM sodium phosphate buffer, pH 7.5, 5 mM dithiothreitol (DTT), and 0.02% NaN$_3$] at 4 °C. The protein concentration of N–SH2 was determined by a Bradford assay kit (Thermo Scientific) due to the lack of Trp residues; the concentration of the other constructs was estimated from UV absorbance at 280 nm with extinction coefficients calculated by ExPASy ProtParam[43] (tSH2: 34380, tSH2$_{PM}$: 32890, tSH2$_{FX}$: 27390, N-SH2 + IA: 17 420 M^{-1} cm^{-1}). The purity of samples was >95% based on SDS/PAGE analysis. The protein samples were 0.8–1.0 mM for relaxation experiments and ∼0.3 mM for titration experiments.

The three peptides, ITP [Ac-PD(pY)EPIRKGQRDL(pY)SGLNQR-NH$_2$], N-IHP [Ac-PD(pY)EPIRKG-NH$_2$] and C-IHP [Ac-QRDL(pY)-SGLNQR-NH$_2$], were purchased from EZBiolab. The purity of all peptides was >95% based on mass spectrometry. The peptides were dissolved in pure water (with pH adjusted to 7) and extensively dialyzed against pure water. The amount of peptide was estimated from UV absorbance at 267 nm with extinction coefficients from the literature:[13] 1304 M^{-1} cm^{-1} for ITP, 652 M^{-1} cm^{-1} for N-IHP and C-IHP. Samples were aliquoted, lyophilized, and stored at −20 °C before use.

NMR data collection

The NMR data were collected on a Bruker Avance-III-800 equipped with a 5 mm TXI Z-gradient probe at 298 K. The T_1, T_2, and NOE ^{15}N relaxation experiments were carried out using the TROSY-based pulse program TRT1ETF3GPSI, TRT2ETF3GP-SI3D, and TRNOEF3GPSI.[44] Relaxation delays were 0.1, 0.1, 0.3, 0.5, 0.5, 0.7, 0.9, 0.9, 1.1, 1.3, 1.3, 1.5 s for T_1 experiments, and 17.3, 17.3, 34.6, 51.8, 51.8, 69.1, 86.4, 86.4, 104, 121, 121, 138 ms for T_2 experiments (the duplicated points were used to estimate uncertainties); the experiments were performed with these delays in a scrambled manner to minimize systematic error. In the ^{1}H^{15}N steady-state NOE experiment, a 3 s recycling time plus a 3 s proton saturation pulse were used for the NOE

spectrum, while a 6 s recycling time was used for the reference spectrum. For titration experiments, a protein sample was sequentially mixed with the aliquoted and lyophilized peptide until >95% saturation, and the chemical shifts were monitored by 2D ^{15}N^{1}H TROSY HSQC experiments with 2048 points in the ^{1}H dimension and 256 points in the ^{15}N dimension. All NMR experiments were performed at least two times.

NMR data analysis

The raw spectra were processed by NMRPipe[45] and visualized by SPARKY 3.113 (T. D. Goddard and D. G. Kneller, University of California, San Francisco, CA). Peaks of ligand-free proteins were assigned according to our previously published data.[18] For peaks in the bound proteins, fast-exchange peaks were assigned by directly tracking the changes of chemical shifts while slow exchange peaks were assigned by TROSY-based 3D HNCA, HNCACB, HN(CO)CA, HN(CO)CACB experiments.

T_1 and T_2 values of individual peaks were extracted by fitting peak intensities $I(t)$ *versus* relaxation delays t to the exponential decay curve: $I(t) = I(0) \times \exp(-t/T)$ (T: T_1 or T_2 value of the peak) using program Sparky2rate (J. P. Loria, Yale University) and Curvefit (A. G. Palmer, Columbia University). NOE values of individual peaks were calculated as the ratio of peak intensities with and without proton saturation. To effectively compare the global motions, the interference of internal motions must be excluded; thus only residues which are within secondary structure elements,[29] with NOE values >0.65,[46] and free from conformational exchange[46] were used for the summary and τ_c fitting. The rotational diffusion correlation time τ_c was fitted by TENSOR 2.0[29] using T_1, T_2, NOE values and the crystal structures of individual domains (PDB ID 1A81). The predicted correlation time for tSH2 and isolated SH2 domains was obtained by HYDRONMR[30] using the crystal structure (PDB ID 1A81, atomic element radius a = 1.5 Å, 298 K).

The model development and fitting procedures of NMR titration-curve and line-shape analyses are described in ESI.†

Acknowledgements

References

1 R. Geahlen, *Biochim. Biophys. Acta, Mol. Cell Res.*, 2009, **1793**, 1115–1127.

2 P. J. Coopman and S. C. Mueller, *Cancer Lett.*, 2006, **241**, 159–173.

3 P. Ruzza, B. Biondi and A. Calderan, *Expert Opin. Ther. Pat.*, 2009, **19**, 1361–1376.

4 P. J. Coopman, M. T. Do, M. Barth, E. T. Bowden, A. J. Hayes, E. Basyuk, J. K. Blancato, P. R. Vezza, S. W. McLeskey, P. H. Mangeat and S. C. Mueller, *Nature*, 2000, **406**, 742–747.

5 Z. A. Stewart and J. A. Pietenpol, *Breast Cancer Res.*, 2001, **3**, 5–7.

6 R. Larive, S. Urbach, J. Poncet, P. Jouin, G. Mascré, A. Sahuquet, P. Mangeat, P. Coopman and N. Bettache, *Oncogene*, 2009, **28**, 2337–2347.

7 J. C. Cambier, *Immunol. Today*, 1995, **16**, 110.

8 M. Reth, *Nature*, 1989, **338**, 383–384.

9 R. L. Geahlen, *Trends Pharmacol. Sci.*, 2014, **35**, 414–422.

10 B. S. Gaul, M. L. Harrison, R. L. Geahlen, R. A. Burton and C. B. Post, *J. Biol. Chem.*, 2000, **275**, 16174–16182.

11 S. Kumaran, R. A. Grucza and G. Waksman, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14828–14833.

12 J. Y. Bu, A. S. Shaw and A. C. Chan, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 5106–5110.

13 R. A. Grucza, J. M. Bradshaw, V. Mitaxov and G. Waksman, *Biochemistry*, 2000, **39**, 10072–10081.

14 T. Chen, B. Repetto, R. Chizzonite, C. Pullar, C. Burghardt, E. Dharm, Z. Zhao, R. Carroll, P. Nunes, M. Basu, W. Danho, M. Visnick, J. Kochan, D. Waugh and A. M. Gilfillan, *J. Biol. Chem.*, 1996, **271**, 25308–25315.

15 E. A. Ottinger, M. C. Botfield and S. E. Shoelson, *J. Biol. Chem.*, 1998, **273**, 729–735.

16 N. J. de Mol, M. I. Catalina, F. J. Dekker, M. J. E. Fischer, A. J. R. Heck and R. M. J. Liskamp, *ChemBioChem*, 2005, **6**, 2261–2270.

17 K. Fütterer, J. Wong, R. A. Grucza, A. C. Chan and G. Waksman, *J. Mol. Biol.*, 1998, **281**, 523–537.

18 Y. Zhang, H. Oh, R. Burton, J. Burgner, R. Geahlen and C. Post, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 11760–11765.

19 R. H. A. Folmer, S. Geschwindner and Y. Xue, *Biochemistry*, 2002, **41**, 14176–14184.

20 U. Grädler, D. Schwarz, V. Dresing, D. Musil, J. Bomke, M. Frech, H. Greiner, S. Jäkel, T. Rysiok, D. Müller-Pompalla and A. Wegener, *J. Mol. Biol.*, 2013, **425**, 309–333.

21 L. M. Keshvara, C. Isaacson, M. L. Harrison and R. L. Geahlen, *J. Biol. Chem.*, 1997, **272**, 10377–10381.

22 D. G. Stupack, E. Li, S. A. Silletti, J. A. Kehler, R. L. Geahlen, K. Hahn, G. R. Nemerow and D. A. Cheresh, *J. Cell Biol.*, 1999, **144**, 777–788.

23 D. Zyss, P. Montcourrier, B. Vidal, C. Anguille, F. Mérezègue, A. Sahuquet, P. H. Mangeat and P. J. Coopman, *Cancer Res.*, 2005, **65**, 10872–10880.

24 L. N. Johnson and R. J. Lewis, *Chem. Rev.*, 2001, **101**, 2209–2242.

25 S.-H. Bae, H. J. Dyson and P. E. Wright, *J. Am. Chem. Soc.*, 2009, **131**, 6814–6821.

26 V. Rolli, M. Gallwitz, T. Wossning, A. Flemming, W. W. A. Schamel, C. Zürn and M. Reth, *Mol. Cell*, 2002, **10**, 1057–1069.

27 E. N. Neumeister, Y. Zhu, S. Richard, C. Terhorst, A. C. Chan and A. S. Shaw, *Mol. Cell. Biol.*, 1995, **15**, 3171–3178.

28 T. Yuwen, C. B. Post and N. R. Skrynnikov, *J. Biomol. NMR*, 2011, **51**, 131–150.

29 P. Dosset, J. C. Hus, M. Blackledge and D. Marion, *J. Biomol. NMR*, 2000, **16**, 23–28.

30 J. García de la Torre, M. Huertas and B. Carrasco, *J. Magn. Reson.*, 2000, **147**, 138–146.

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

This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 5807–5818 | 5817

32 D. M. Crothers, P. E. Cole, C. W. Hilbers and R. G. Shulman, *J. Mol. Biol.*, 1974, **87**, 63–88.

33 B. D. Rao, *Methods Enzymol.*, 1989, **176**, 279–311.

34 H. M. McConnell, *J. Chem. Phys.*, 1958, **28**, 430.

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

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

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

38 J. E. Ladbury, M. A. Lemmon, M. Zhou, J. Green, M. C. Botfield and J. Schlessinger, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 3199–3203.

39 Y. Zhou and R. Abagyan, *Folding Des.*, 1998, **3**, 513–522.

40 W. Gan and B. Roux, *Proteins*, 2009, **74**, 996–1007.

41 M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1–16.

42 C. E. Hughes, A. Y. Pollitt, J. Mori, J. A. Eble, M. G. Tomlinson, J. H. Hartwig, C. A. O'Callaghan, K. Futterer and S. P. Watson, *Blood*, 2010, **115**, 2947–2955.

43 M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser, *Methods Mol. Biol.*, 1999, **112**, 531–552.

44 G. Zhu, Y. Xia, L. K. Nicholson and K. H. Sze, *J. Magn. Reson.*, 2000, **143**, 423–426.

45 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.

46 N. Tjandra, S. E. Feller, R. W. Pastor and A. Bax, *J. Am. Chem. Soc.*, 1995, **117**, 12562–12566.

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).
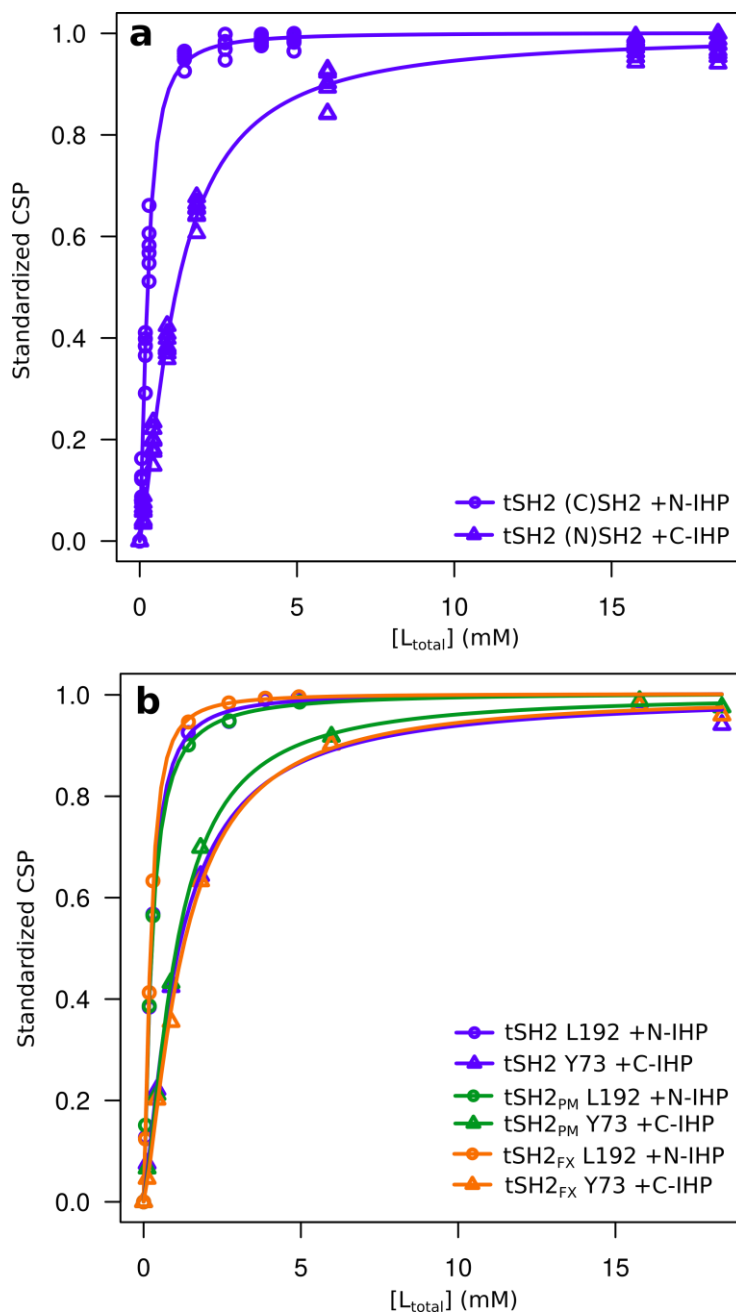
Fig. S1 (**a**) standardized CSP (chemical shift perturbation) titration plot for resonances from residues in each domain of tSH2 (0.3 mM) upon addition of the IHP peptides. These residues correspond to the contacts of tSH2 with dp-ITAM (PDB ID 1A81). Residues within the same domain showed similar behavior 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.

3

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,

$$K_N = \frac{[\text{NC}][\text{L}]}{[\text{N}^\text{L}\text{C}]} = \frac{[\text{NC}^\text{L}][\text{L}]}{[\text{N}^\text{L}\text{C}^\text{L}]} \tag{1}$$

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

$$K_C = \frac{[\text{NC}][\text{L}]}{[\text{NC}^\text{L}]} = \frac{[\text{N}^\text{L}\text{C}][\text{L}]}{[\text{N}^\text{L}\text{C}^\text{L}]} \tag{2}$$

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

$$[\text{P}_\text{total}] = [\text{NC}] + [\text{N}^\text{L}\text{C}] + [\text{NC}^\text{L}] + [\text{N}^\text{L}\text{C}^\text{L}] \tag{3}$$

$$[\text{L}_\text{total}] = [\text{L}] + [\text{N}^\text{L}\text{C}] + [\text{NC}^\text{L}] + 2[\text{N}^\text{L}\text{C}^\text{L}] \tag{4}$$

From eqn (1-4), the relationship between [L], $[\text{P}_\text{total}]$, $[\text{L}_\text{total}]$, K_N and K_C is

$$[\text{L}]^3 + [\text{L}]^2(2[\text{P}_\text{total}] - [\text{L}_\text{total}] + K_N + K_C)$$
$$+ [\text{L}]\big((K_N + K_C)[\text{P}_\text{total}] - (K_N + K_C)[\text{L}_\text{total}] + K_N K_C\big) - [\text{L}_\text{total}]K_N K_C = 0 \tag{5}$$

With given values of $[\text{P}_\text{total}]$, $[\text{L}_\text{total}]$, 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):

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

$$[\text{N}^\text{L}\text{C}] = \frac{[\text{L}][\text{NC}]}{K_N} \tag{7}$$

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

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

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

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

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

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

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

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

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 $[P_{total}]$, $[L_{total}]$, 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).
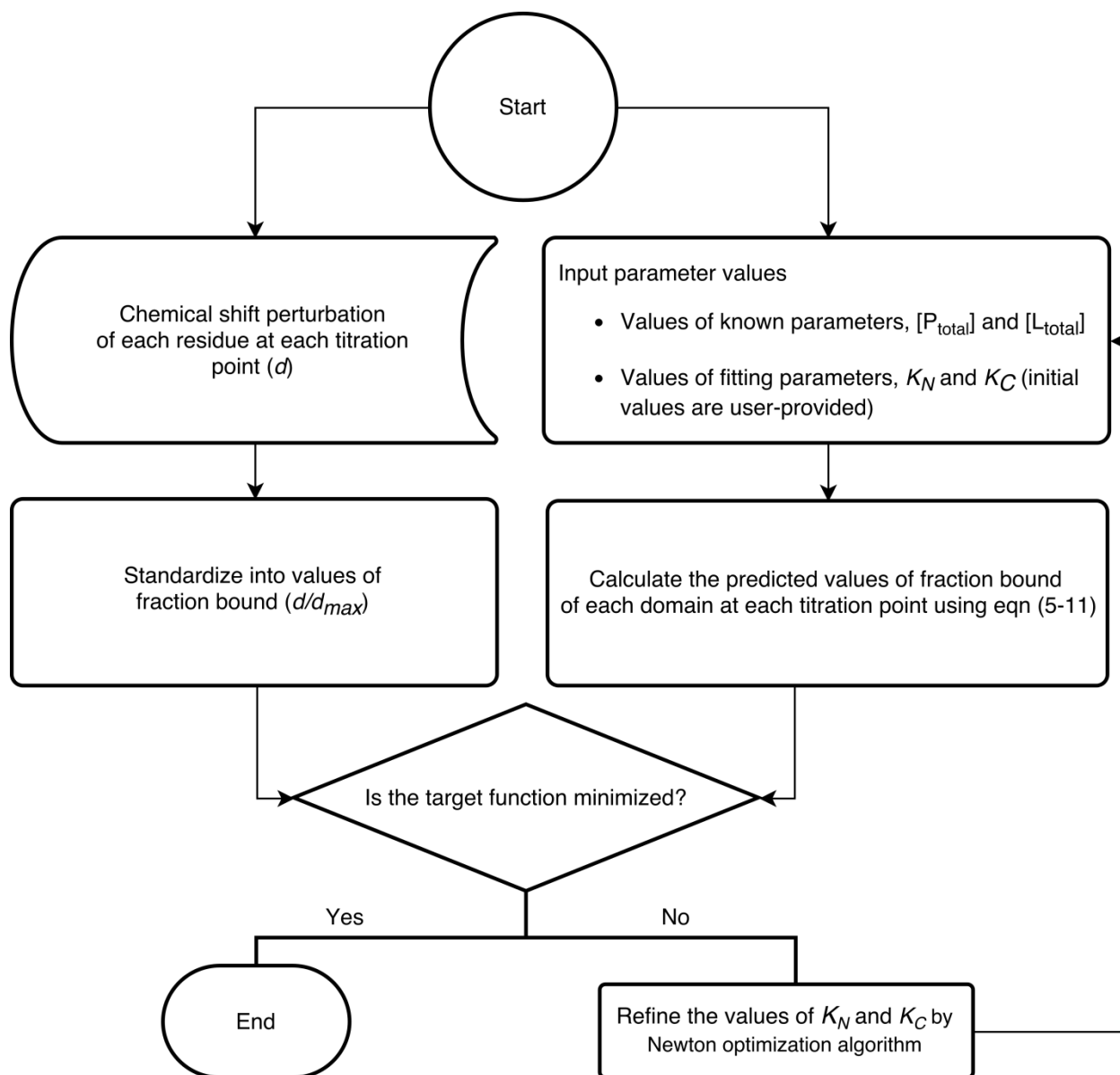
Fig. S2 Flow chart of the titration curve analysis for tSH2/IHP complexes with the 4-state model. The target function to be minimized is the sum of squared errors between the experimentally and computationally determined fraction-bound values of each residue at each titration point.
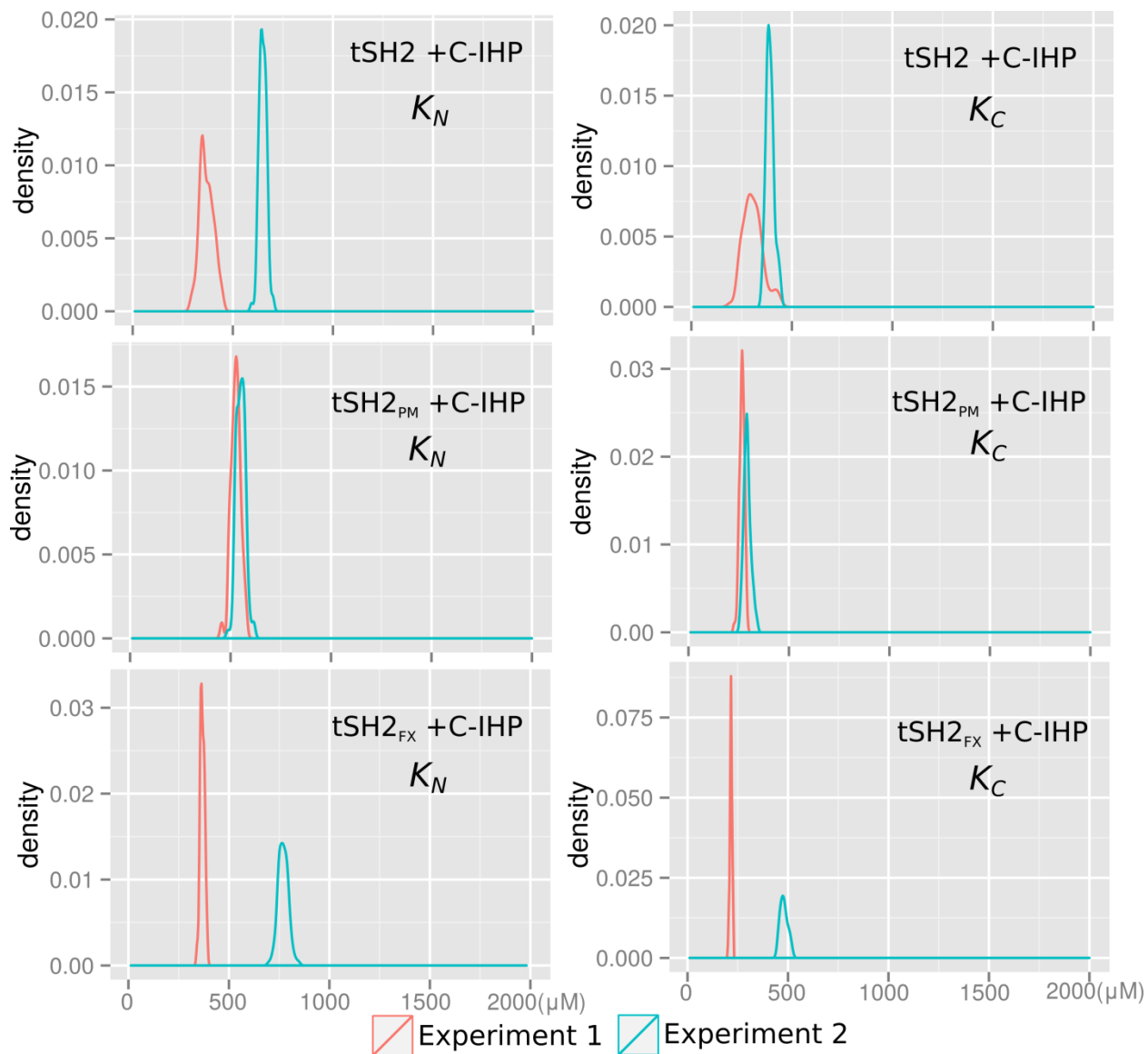
Fig.S3 The distribution of fitted values of dissociation constants of C-IHP binding to the three tSH2 constructs using the **titration-curve analysis** and **4-state model**. Two independent experiments were performed for each complex and 100 Monte-Carlo simulations were performed for each experimental 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:

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

where

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

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

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

$\mathbf{P}$ is a vector of equilibrium concentrations of the four species calculated by eqn (5-9) at the given titration point. $\mathbf{R}$ is a diagonal matrix with elements equal to the transverse relaxation rates of the resonance of the nucleus in the four species in $\mathbf{P}$. $\mathbf{\Omega}$ is a diagonal matrix for the intrinsic resonant frequency (ω_0) of the given nucleus in each of the four species in $\mathbf{P}$. $\mathbf{R}$ and $\mathbf{\Omega}$ 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 $2R_{2,0}$). $\mathbf{K}$ is the 4 by 4 exchange matrix

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

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 [P_{total}], [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):

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

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).
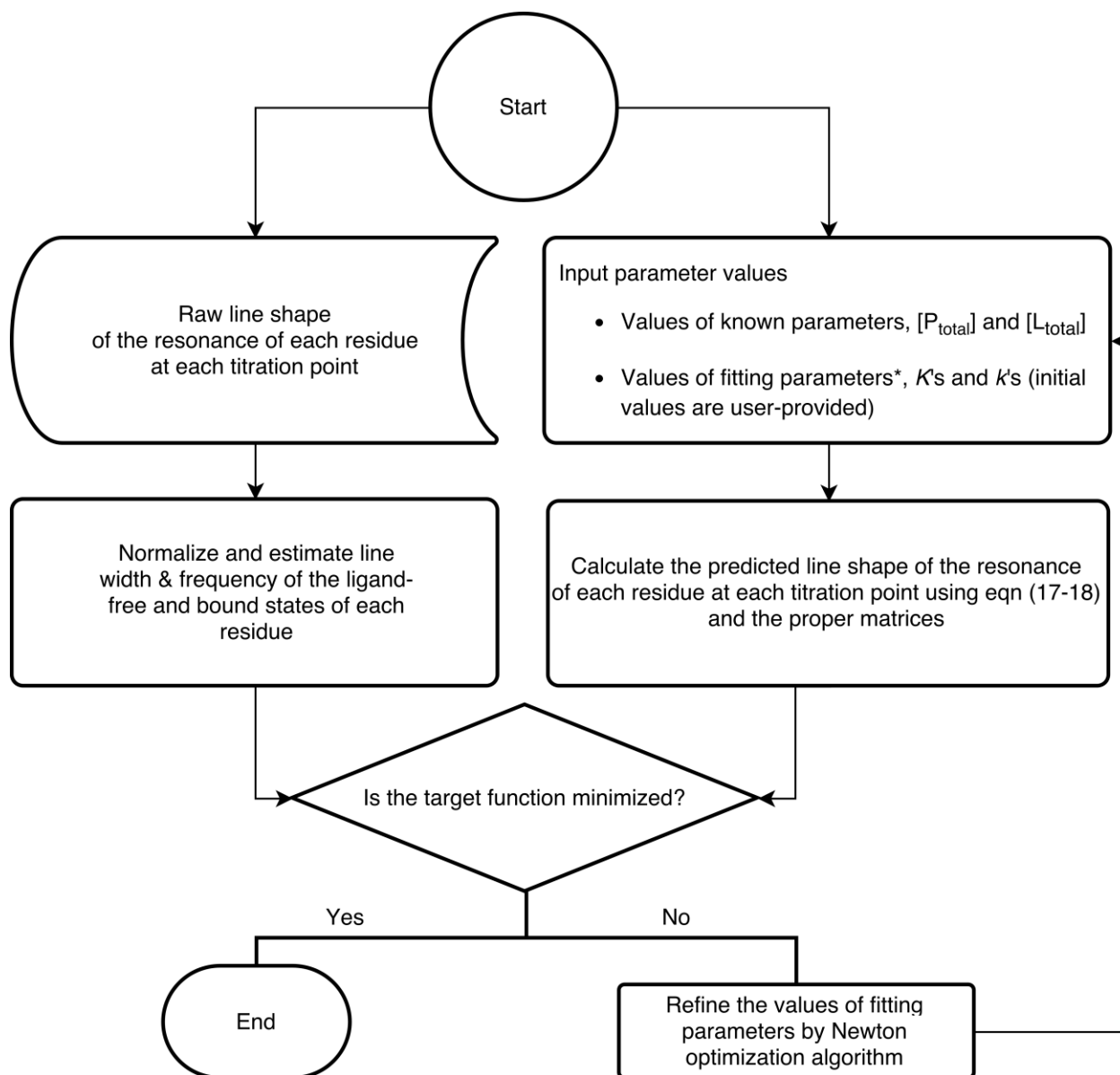
Fig.S4 Flow chart of the line-shape analysis procedures. The target function to be minimized is the sum of squared errors between the experimentally and computationally determined line-shape data of each residue at each titration point. The fitting parameters vary for different models, while the other parameters in the models are fixed, where appropriate, using known values.
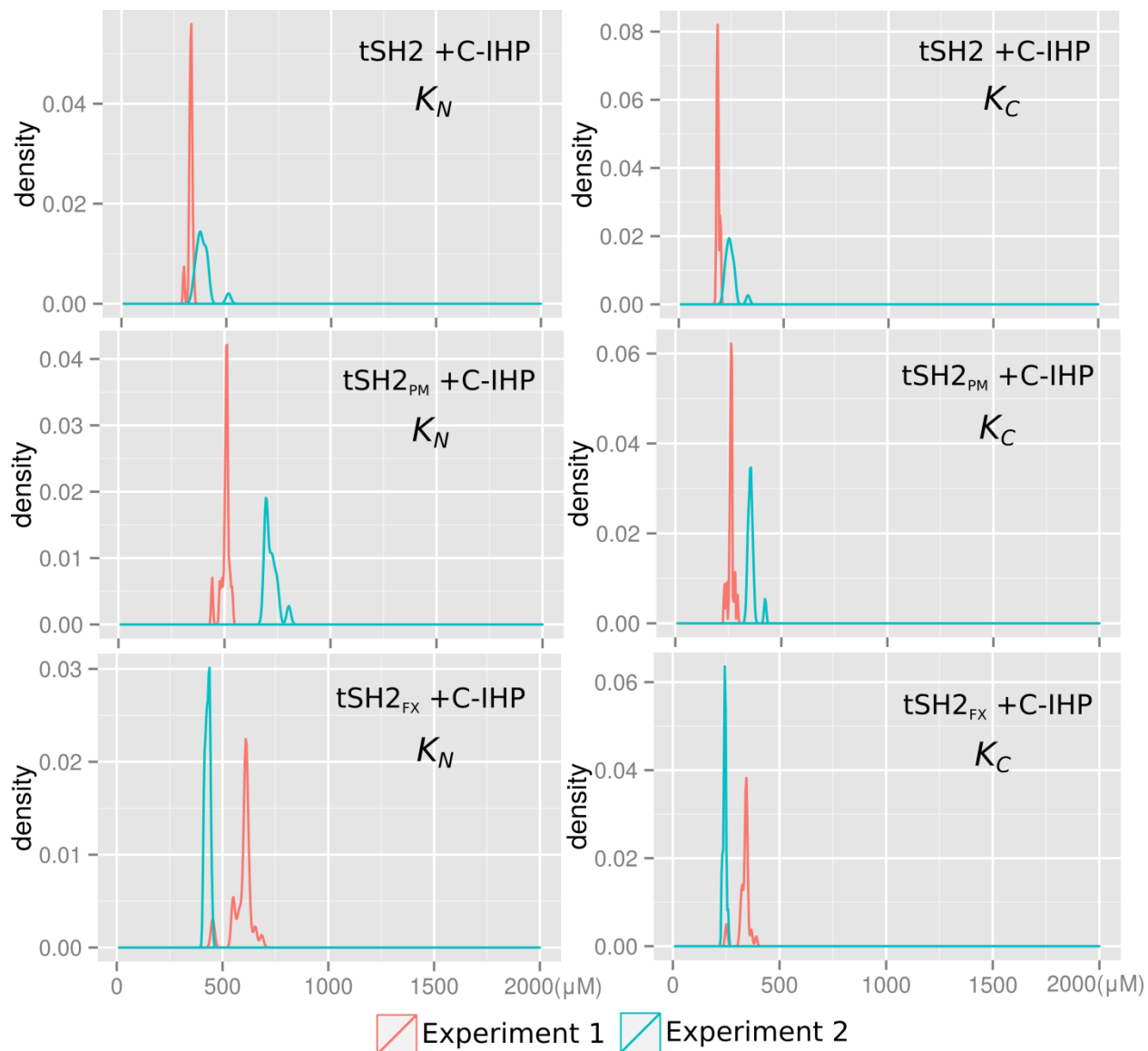
Fig.S5 The distribution of fitted values of dissociation constants of C-IHP binding to the three tSH2 constructs using the **line-shape analysis** and **4-state model**. Two independent experiments were performed for each complex and 100 bootstrapping simulations were performed for each experimental 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.
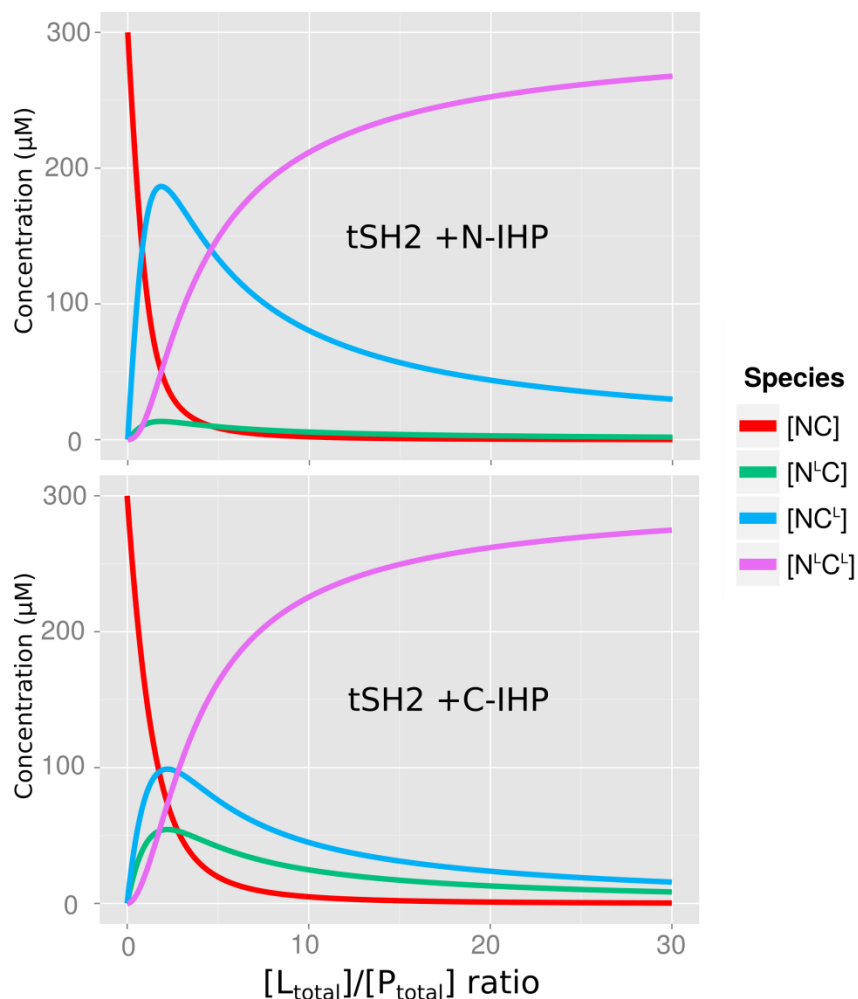
Fig.S6 The simulated change of concentration of each species for the tSH2 construct binding with IHPs in the 4-state binding model. [P_{total}] is 0.3 mM similar as that for NMR experiments. The simulated curves are similar for all three tSH2 constructs, because the binding affinity of a specific SH2 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 :

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

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

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

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

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

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

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

With given values of $[P_{total}]$, $[L_{total}]$, 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):

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

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

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

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:

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

$$k'^{N}_{on} = \frac{k'^{N}_{off}}{K'_{N}}$$ (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:

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

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

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

$$\mathbf{K} = \begin{pmatrix} -k^{CN}_{on}[L] & k^{CN}_{off} & 0 \\ k^{CN}_{on}[L] & -k^{CN}_{off} - k'^{N}_{on} & k'^{N}_{off} \\ 0 & k'^{N}_{on} & -k'^{N}_{off} \end{pmatrix}$$ (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.
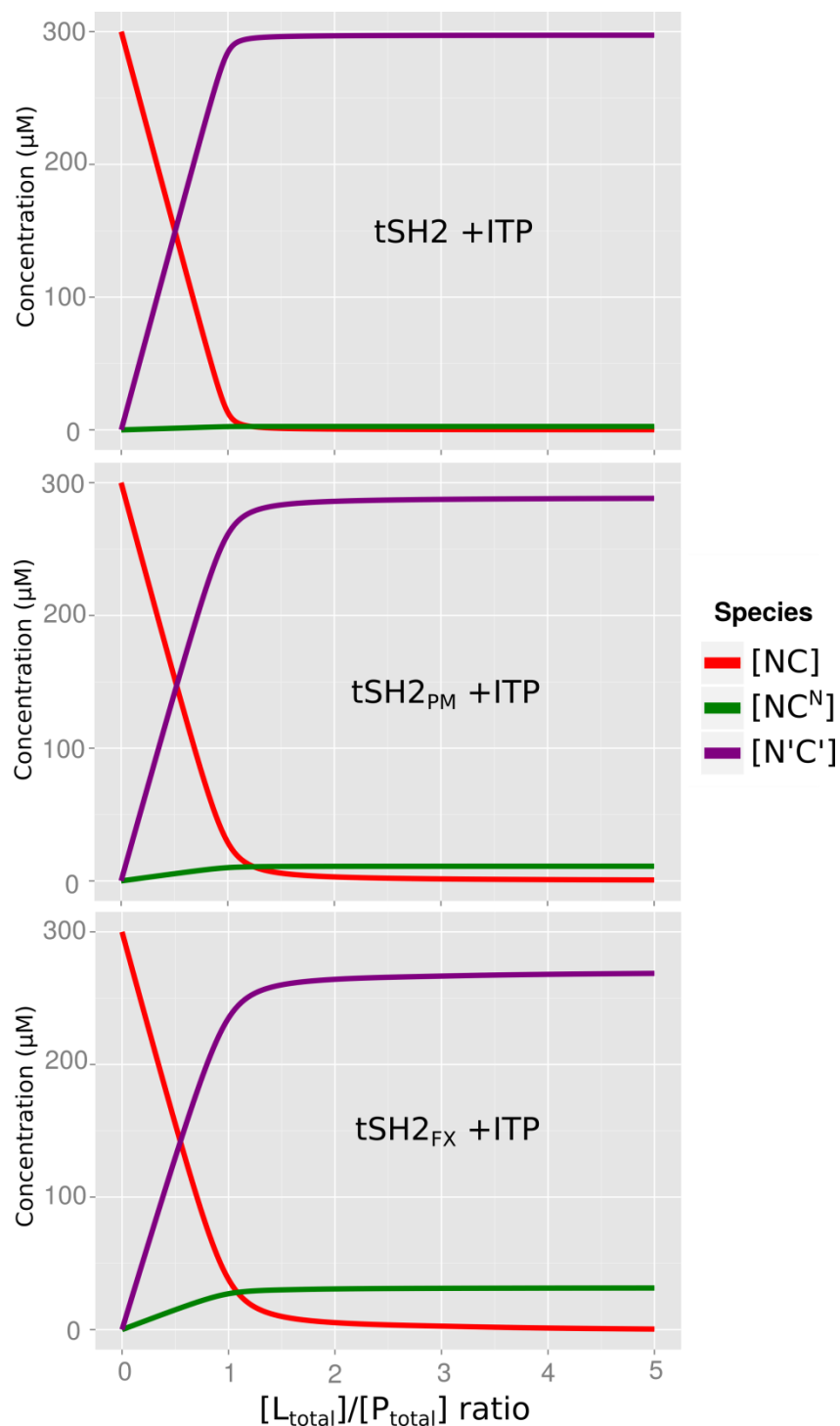
Fig.S7 The simulated change of concentration of each species for the three tSH2 constructs binding with ITP in the 3-state binding model. $[P_{total}]$ is 0.3 mM similar as that for NMR experiments. For each complex, the curves were simulated using the fitted intra-molecular isomerization constant (K'_N) 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:

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

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

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

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

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

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

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

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

$$[P_{total}] = [NC] + [N^N C] + [N^C C] + [NC^N] + [NC^C] + [N'C'] \\ + [N^N C^N] + [N^C C^N] + [N^N C^C] + [N^C C^C] \tag{42}$$

$$[L_{total}] = [L] + [N^N C] + [N^C C] + [NC^N] + [NC^C] + [N'C'] \\ + 2[N^N C^N] + 2[N^C C^N] + 2[N^N C^C] + 2[N^C C^C] \tag{43}$$

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

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

where

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

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

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

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

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):

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

$$[N^N C] = \frac{[L][NC]}{K_{NN}} \tag{50}$$

$$[N^C C] = \frac{[L][NC]}{K_{NC}} \tag{51}$$

$$[NC^N] = \frac{[L][NC]}{K_{CN}} \tag{52}$$

$$[NC^C] = \frac{[L][NC]}{K_{CC}} \tag{53}$$

$$[N'C'] = \frac{[L][NC]}{K'_N K_{CN}} \tag{54}$$

$$[N^N C^N] = \frac{[L]^2[NC]}{K_{CN}K_{NN}} \tag{55}$$

$$[N^C C^N] = \frac{[L]^2[NC]}{K_{CN}K_{NC}} \tag{56}$$

$$[N^N C^C] = \frac{[L]^2[NC]}{K_{CC}K_{NN}} \tag{57}$$

$$[N^C C^C] = \frac{[L]^2[NC]}{K_{CC}K_{NC}} \tag{58}$$

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:

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

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

$\mathbf{E}$ is a 10 by 10 identity matrix. $\mathbf{P}$ is a vector of equilibrium concentrations of the ten species calculated from eqn (44-58) for the given titration point. $\mathbf{R}$ and $\mathbf{\Omega}$ 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). $\mathbf{K}$ is the 10 by 10 exchange matrix

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

Where

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

$$a_{2,1} = k_{on}^{NN}[L] \tag{69}$$

$$a_{3,1} = k_{on}^{NC}[L] \tag{70}$$

$$a_{4,1} = k_{on}^{CN}[L] \tag{71}$$

$$a_{5,1} = k_{on}^{CC}[L] \tag{72}$$

$$a_{1,2} = k_{off}^{NN} \tag{73}$$

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

$$a_{7,2} = k_{on}^{CN}[L] \tag{75}$$

$$a_{9,2} = k_{on}^{CC}[L] \tag{76}$$

$$a_{1,3} = k_{off}^{NC} \tag{77}$$

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

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

$$a_{8,3} = k_{on}^{CN}[L] \tag{80}$$

$$a_{10,3} = k_{on}^{CC}[L] \tag{81}$$

$$a_{1,4} = k_{off}^{CN} \tag{82}$$

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

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

$$a_{7,4} = k_{on}^{NN}[L] \tag{85}$$

$$a_{8,4} = k_{on}^{NC}[L] \tag{86}$$

$$a_{1,5} = k_{off}^{CC} \tag{87}$$

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

$$a_{9,5} = k_{on}^{NN}[L] \tag{89}$$

$$a_{10,5} = k_{on}^{NC}[\text{L}] \tag{90}$$

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

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

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

$$a_{2,7} = k_{off}^{CN} \tag{94}$$

$$a_{4,7} = k_{off}^{NN} \tag{95}$$

$$a_{7,7} = -k_{off}^{CN} - k_{off}^{NN} \tag{96}$$

$$a_{3,8} = k_{off}^{CN} \tag{97}$$

$$a_{4,8} = k_{off}^{NC} \tag{98}$$

$$a_{8,8} = -k_{off}^{CN} - k_{off}^{NC} \tag{99}$$

$$a_{2,9} = k_{off}^{CC} \tag{100}$$

$$a_{5,9} = k_{off}^{NN} \tag{101}$$

$$a_{9,9} = -k_{off}^{CC} - k_{off}^{NN} \tag{102}$$

$$a_{3,10} = k_{off}^{CC} \tag{103}$$

$$a_{5,10} = k_{off}^{NC} \tag{104}$$

$$a_{10,10} = -k_{off}^{CC} - k_{off}^{NC} \tag{105}$$

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.
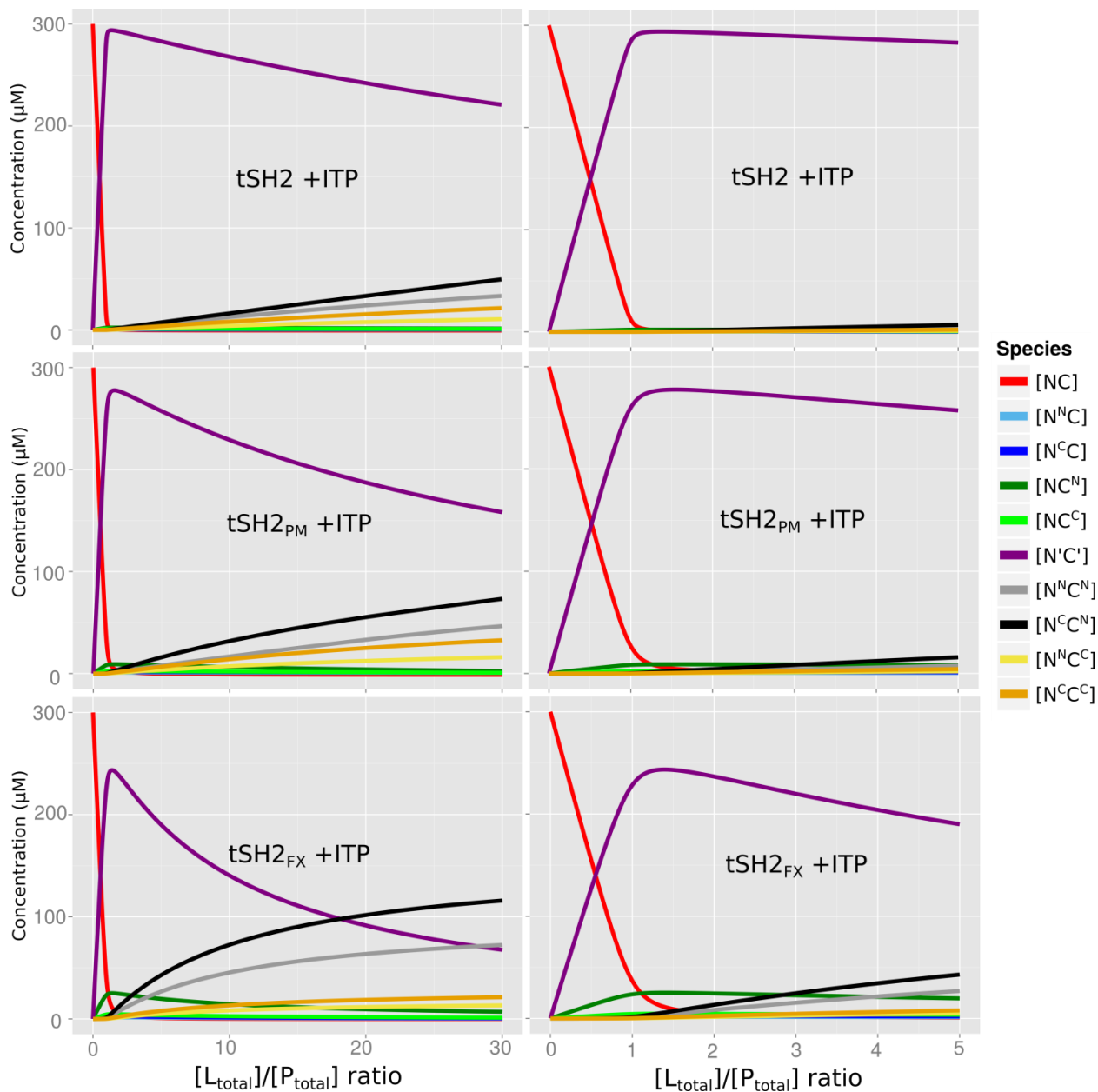
Fig.S8 The simulated change of concentration of each species for the tSH2 constructs binding with ITP in the 10-state binding model, with two ranges of the $[L_{total}]/[P_{total}]$ ratio for each construct. $[P_{total}]$ is 0.3 mM similar as that for NMR experiments. For each construct, the curves were 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.