Multidomain assembled states of Hck tyrosine kinase in solution

Sichun Yang,a Lydia Blachowicz,a Lee Makowski,b and Benoît Roux*a,†

Abstract

An approach combining small-angle X-ray solution scattering (SAXS) data with coarse-grained (CG) simulations is developed to characterize the assembly states of Hck, a member of the Src-family kinases, under various conditions in solution. First, a basis set comprising a small number of assembly states is generated from extensive CG simulations. Second, a theoretical SAXS profile for each state in the basis set is computed by using the Fast-SAXS method. Finally, the relative population of the different assembly states is determined via a Bayesian-based Monte Carlo procedure. The study establishes the concept of assembly states with experimental data. The results indicate that multiple assembly states are generally present, and that the population equilibrium among these states responds to both mutations and the presence of signaling peptides binding to the SH2 or SH3 domains.

Author contributions: S.Y. and B.R. designed research; S.Y. performed research; S.Y. and L.M. analyzed data; and S.Y., L.M., L.B., and B.R. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.
†To whom correspondence should be addressed. E-mail: roux@uchicago.edu.
structures, (ii) the topological connectivity of the polypeptide chain via flexible linkers, and (iii) the excluded volume from the residue-residue core repulsion. The CG model also accounts for the intramolecular interactions involved in the stabilization of the down-regulated assembled form (3): the tyrosine-containing C-terminal “Ctail” of the kinase domain with the SH2 domain, and the proline SH2-kinase “linker” with the SH3 domain. As described in SI Appendix, both interactions were incorporated in the CG model, and adjusted to match the $K_d$ from experiments for the SH2-binding Ctail peptide, and the SH3-binding proline peptide. The active and inactive conformations of the catalytic domain are also included in the multistate model as described previously (16, 17) to account for the ability of full-length Hck to adopt different conformations. To ensure proper sampling of various assembly conformations, simulations using different initial conditions were conducted, ranging from fully assembled to disassembled states, from increased to decreased binding interactions, and from destabilizing stable configurations to lowered transition barriers.

A large number of configurations were generated using MD simulations based on the CG model. A two-step clustering scheme was used to organize the large amount of information generated by the simulations of the CG model into a manageable form. Initially, 25 structural clusters were determined from the trajectories from a structural clustering based on residue-residue distances similarity criterion (16, 17). More detail on the clustering is given in SI Appendix: Fig. S1. Then, average scattering patterns were calculated for members of each cluster using the Fast-SAXS method (18), and the clusters were further regrouped into nine assembly states on the basis of a scattering pattern similarity criterion. The relationship between the nine states according to this metric is shown diagrammatically in SI Appendix: Fig. S1, and a representative conformation from each state is shown in Fig. 1. It should be noted that the configurations within those states are regrouped according to the similarity of their scattering patterns, and that they may display conformational differences. This set of nine assembly states, extracted from the simplified CG simulations, is then used as a basis-set for extracting the multidomain organization of Hck in solution from the experimentally observed SAXS patterns.

**Bayesian-Based Monte Carlo Analysis of SAXS Data.** The calculated profile $I_{\text{calc}}$ is an average over the $N_s = 9$ states, weighted by the fractional population $P_i$ of state $i$,

$$I_{\text{calc}}(q) = \sum_{i=1}^{N_s} P_i I_i(q),$$

where $I_i(q)$ is the theoretical scattering profile of state $i$ (see examples in SI Appendix: Fig. S2). The goal of the analysis is to determine the set of $N_s$ fractional population $\{P_i\}$ from the SAXS data. To this end, we define a dimensionless scoring function, $\chi^2([P_i])$, measuring the agreement between the calculated and experimental SAXS profiles,

$$\chi^2([P_i]) = \sum_{q_{\text{min}}}^{q_{\text{max}}} \frac{1}{\sigma(q)} (\log I_{\text{calc}}(q) - \log I_{\text{exp}}(q) - \Delta)^2,$$

where $q_{\text{min}}$ and $q_{\text{max}}$ are the lower and upper limit of $q$-range of the experimental profile of $I_{\text{exp}}(q)$, respectively, and $\Delta$ is the offset between $\log I_{\text{calc}}$ and $\log I_{\text{exp}}(q)$ at $q = q_{\text{min}}$. The estimated experimental SAXS uncertainty, $\sigma(q)$, is assumed to be $[\delta \log I_{\text{exp}}(q)]^{1/2}$. Assuming independent experimental errors and a uniform prior, the probability distribution $\exp(-\chi^2)$ is sampled in the space of the model parameters, $\{P_i\}$, according to a Bayesian-Based Monte Carlo (BMC) algorithm. The procedure yields the fractional population $P_i$ for the $N_s$ states, along with their uncertainties extracted from the fluctuations.

**BSS-SAXS Characterization of Hck Assembly.** SAXS data were collected for the wild-type Hck, and for the high-affinity Ctail mutant (Hck-YEEI), in the presence of two types of external peptide ligands: (p2) a high-affinity phosphorylated tyrosine-containing SH2-binding peptide (6); (p3) a high-affinity SH3-binding PPII peptide (21 and 22). All the experimental scattering profiles are shown in Figs. 2, 3. For the sake of completeness, a simple Guinier analysis was performed for all the SAXS data shown in Figs. 2 and 3, and the radii of gyration $R_g$ were extracted by fitting the scattering patterns in the range up to 0.05 Å$^{-1}$. The results are reported in the Figure captions.
The experimental SAXS data from wild-type Hck in the absence of external peptides is shown in Fig. 2A. It is observed that the SAXS profile calculated from the population derived from the BMC analysis (red) accurately reproduces the experimental data (black). The fit is achieved with the population of assembly states determined via the BMC analysis. The latter indicates that the enzyme is predominantly in the assembled state 1, with the population of 82%. The more weakly populated states include the fully disassembled state 6 and the partially disassembled state 8 with SH2-dissociated but SH3-bound similar to the partially disassembled crystal structure of c-Src (10). As shown in Fig. 2A, this combination of states best fits the SAXS data. The experimental SAXS data from the high-affinity Ctail mutant YEEI is shown in Fig. 2B. In this case too, the assembled state 1 dominates the three-dimensional organization, with the population of 83%.

To investigate how ligand binding affects the equilibrium assembly of the multidomain complex, SAXS data were also collected from Hck with two types of peptide ligands. As shown in Fig. 3A, there is a substantial shift of wild-type Hck from a compact state to disassembled states in the presence of 250 μM of the high-affinity phosphorylated SH2-binding peptide p2. Comparison of Figs. 2A and 3A shows that the population of the compact state 1 decreases from 82% to 22%, while that of the disassembled state 6 increases from 8% to 29%. Repeating this experiment for the Hck-YEEI Ctail mutant yields a very different result. The BMC analysis given in Fig. 3B indicates that, for the Hck-YEEI mutant, the population of the assembled state is reduced only to 62% upon addition of the peptide p2. Here, the disassembled and compact state 6, where SH2 and SH3 are disassociated but the overall Hck architecture remains compact, is also populated.

The effect of the SH3-binding peptide p3 is shown in Fig. 3C. BMC analysis of the SAXS data indicates that p3 destabilizes the assembled down-regulated state 1 and promotes the disassembled compact state 6. Comparison of Figs. 2A and 3C shows that the population of state 1 decreases from 82% to 7% and that of state 6 increases from 8% to 50% in the presence of peptide p3. Also, the population of the SH3-displaced states 2 and 3 increases, reaching a total combined population of about 39%. Cross-correlation analysis of the population of states 2 and 3, however, suggests that they are highly correlated and that determining their relative weight is at the limit of resolution of the present BMC analysis.

Lastly, the spatial organization of Hck was examined in the presence of both the p2 and p3 peptides simultaneously. The results of the BMC analysis of SAXS data, shown in Fig. 3D, demonstrate that any remnant of the assembled down-regulated state is wiped out upon the binding of p2 and p3. Here, the fully disassembled states 5 and 6 are representative of the spatial organization. Both states are disassembled when Ctail is released from SH2 and SH3 is disassociated, but they differ in overall size: state 5 is extended (SI Appendix: Fig. S8) and state 6 exhibits a compact form (SI Appendix: Fig. S9). Consistent with the BSS-SAXS analysis, \( R_g \) reaches its largest value (31.7 Å) under these conditions.

**Discussion**

The SAXS data is consistent with an increasing disassembled state for Hck as it is perturbed by the signaling peptides. In
particular, the radius of gyration $R_g$ extracted from the SAXS data via a Guinier analysis is typically smaller in the absence of binding peptides (Fig. 2) than in their presence (Fig. 3). Also consistent with this result, the smallest value of 27.1 Å is observed of binding peptides (Fig. 2) than in their presence (Fig. 3). Also consistent with this result, the smallest value of 27.1 Å is observed.

Fig. 3. The wild-type and mutant Hck in solution in the absence and presence of SH2-binding (p2) and SH3-binding (p3) peptides. In all cases, the scattering pattern (left), the assembly state population (center), and the relative $\chi^2$ scores for different combinations (right) are shown. (A) SAXS data of Hck in the presence of p2 at a concentration of 250 μM (black) and the BMC simulated profile (red) are shown. Addition of p2 causes a shift in the equilibrium toward a coexistence of assembled and various disassembled states. $R_g$ is 29.3 ± 0.2 Å. Relative $\chi^2$ scores for different solutions of the state combination. (B) SAXS data of Hck-YEEI in the presence of p2 at a concentration of 250 μM (black) and the BMC simulated profile (red) are shown. $R_g$ is 28.1 ± 0.1 Å. Furthermore, in contrast to the Hck-YEEI scattering in the absence of p2 (Fig. 2B), these results demonstrate that the addition of p2 to a Hck-YEEI solution causes an increase in population of the disassembled and compact state 6. (C) SAXS data of Hck-YEEI in the presence of p3 peptide at a concentration of 1 mM (black) and the BMC simulated profile (red) are shown. $R_g$ is 29.0 ± 0.3 Å. (D) Addition of both p2 and p3 peptides shifts the equilbrium of Hck conformations toward disassembled states. SAXS data of the wild-type Hck in the presence of both p2 at a concentration of 250 μM and p3 at a concentration of 1 mM (black) and the BMC simulated profile (red). $R_g$ is 31.7 ± 0.2 Å.
for the EEI Ctail mutant of Hck (Fig. 2B), whereas the largest value of 31.7 Å is observed for the wild-type Hck in the presence of peptides p2 and p3 (Fig. 3D). Nevertheless, although a simple analysis of SAXS data based on the variations of $R_g$ is indicative of broad structural changes, the scattering pattern does not, by itself, provide a straightforward characterization of the assembly of a complex multidomain protein such as Hck.

To determine quantitatively how various factors can affect the configurational landscape of Hck, we set out to estimate the relative population of a set of possible assembly states from the experimental SAXS data using a BMC analysis. This analysis leans on the configurations generated by simulating a simplified CG model. The set of nine assembly states extracted from those simulations serves as a basis-set covering the wide range of compact and loose configurations that Hck could possibly adopt. In the BSS-SAXS reconstruction, the set of nine assembly states constitutes a Bayesian prior in the analysis (23). One configuration from each of the assembly states is illustrated in Fig. 1. However, it is emphasized that each state actually comprises a large ensemble of configurations (SI Appendix: Figs. S4–S12). Such conformational heterogeneity is consistent with the existence of a broad continuum of accessible configurations, but it could also reflect our limited ability to resolve the conformation of a multidomain protein unambiguously using only the information provided by the scattering patterns. Without additional experimental data permitting a more detailed structural description, it is best to exert some caution when interpreting the results of the BSS-SAXS analysis in terms of “state” population.

The nine states in Fig. 1 can be resolved from SAXS data, i.e., differences in scattering intensity are larger than experimental uncertainty. Distinguishability of the states included in the set is measured from the uncertainty-weighted difference in average SAXS patterns (1(q)) (see also SI Appendix: Table S1). The set includes a wide range of different configurations, varying in their multidomain organization and overall architecture. State 1 is similar to the X-ray structure of Hck corresponding to the fully assembled down-regulated kinase (3, 4), while state 8 is analogous to the partially activated state observed in one X-ray structure of c-Src (10). Fluctuations are important and there are subtle, though important, differences among some states. For example, states 5 and 6 are both disassembled, but the former is more extended than the latter. State 2 and 3 correspond to SH3-displaced configurations, lacking the direct interaction between the SH3 and the PPII linker. In some disassembled configurations, the SH3 and SH2 domains maintain the relative orientation observed in the assembled form (state 1), consistent with a previous simulation study (24).

The result of the BMC analysis of SAXS data for the wild-type Hck is shown in Fig. 2A. There as well, the assembled state 1 dominates the three-dimensional organization. The lack of increase in the population of assembly state 1 is somewhat surprising because the high-affinity Ctail mutant is known to down-regulate the kinase (6), although a Guinier analysis indicates that $R_g$ shifts from 28.1 Å to 27.1 Å (Fig. 2A and B). As shown in Fig. 3A and B, the presence of the SH2-binding peptide p2 destabilizes the assembled down-regulated state, the effects being more important for the wild-type Hck than for the Hck-EEI Ctail mutant. The population of assembly state 1 is reduced to 22% for the wild-type Hck, while it is reduced only to 62% for the Hck-EEI mutant. This destabilization is also reflected in the $R_g$ of 29.3 Å and 28.1 Å, for the wild-type and Hck-EEI mutant, respectively. The shift toward disassembled configurations is consistent with the increase in kinase activity observed in the presence of SH2-binding ligands (7). Nevertheless, the nonnegligible population of assembly state 1 remaining in the presence of high-affinity SH2-binding peptide is somewhat surprising. The peptide comprises a phosphorylated tyrosine and is present at a nearly saturating concentration of 250 μM. In spite of this saturation, the intramolecular SH2-Ctail interactions for both the Hck-EEI mutant and, to a lesser extent, wild-type Hck compete against the binding of a very high-affinity peptide and protect the assembly state 1. The observation can be rationalized in terms of a simple equilibrium model between an assembled state A and a disassembled state D, with the assumption that only the latter can bind the external ligand. The equilibrium for the three state process, $A = D + L = DL$, is given by $[D]/[A] = p$, and $[DL]/[D] = [L]/K_p$, where $p$ is the natural propensity of occurrence of the D state, and $K_p$ is the dissociation constant, respectively. The probability of the assembled state is then equal to $P_A = (1 + p + p[L]/K_p)^{-1}$. Because $p$ is smaller for the high-affinity Ctail mutant Hck-EEI than for the wild-type Hck, the intramolecular interaction is able to override the strong bimolecular association $(p_{\text{EEI}}[L]/K_p ▶ p_{\text{EEI}}[L]/K_p)$, thus protecting the assembled down-regulated conformation.

The action of the SH3-binding peptide p3 is meant to mimic a signal such as produced by the Nef PxxP motif, which was previously used to induce kinase activation by releasing the SH3 module from the PPII linker (26). The analysis of the SAXS data for the Hck-EEI mutant in the presence of p3 shown in Fig. 3C indicates that the binding of p3 increases the population of assembly states 2 and 3, corresponding to two possible SH3-displaced orientations (see also SI Appendix: Figs. S5 and S6). A number of in vitro and in vivo data including cell based assays and mutagenesis have also supported the concept of such an SH3-displaced model in kinase activation (5, 27). While the present results are clearly at the limit of resolution available from the
analysis of SAXS data, they support the concept of SH3-displaced states.

The classic view of Src activation has typically invoked some loose disassembled conformation, in which all autoinhibitory intramolecular interactions are released. The analysis of the SAXS data for the wild-type Hck mutant in the presence of both the p2 and p3 peptides shown in Fig. 3D yields a representation of such a state of Hck in solution. Of particular interest, the assembled state 1 is wiped out while the most populated states, 5 and 6, differ from those detected under the previous conditions. This result demonstrates that the multidomain protein has access to a wide band of different conformations and that different conditions can shift the population of the states in various directions.

Conclusion

An integrated framework combining SAXS experiments and computational modeling was introduced to quantitatively characterize the assembly of Hck kinase in solution. The BSS-SAXS analysis provides a view of the rich possibilities of three-dimensional conformational organizations accessible to Src in solution, going beyond the down-regulated assembled or fully disassembled states, which may be associated with kinase activation. The method offers a complementary and powerful approach to characterize multidomain molecular assemblies in solution, especially when multiple conformations can coexist. An important finding from the analysis is that multiple assembly states are required to construct a proper basis-set covering the conformational landscape of Hck. Available crystal structures are not enough to match the experimental SAXS data. Of particular interest, a large fraction of the kinase conformations exist in the assembled down-regulated state, even when the Tyr527 in the C terminus of the catalytic domain is not phosphorylated. Furthermore, different external signals lead to different distributions among the possible assembly states. These observations point to a molecular interpretation of the allosteric response of Hck to various biochemical signals that is in accord with the general concept of a shift in equilibrium populations (28, 29).

Methods

Experiments

Wild-type HCK3D or the HCK3D.YEEI mutant were coexpressed in bacteria with Yohp phosphatase and purified as described by Seelig et al. (30). The SH2 binding peptide (referred to as p2; sequence EPQ[pY]EIEIKPKD, where pY is phosphotyrosine) and the SH3 binding peptide (referred to as p3; sequence VSLLARPLPLP) were synthesized in house. SAXS data were collected at the BioCAT-18BD from the Advanced Photon Source. Data were collected using a flow-cell kept at 10 °C with a series of 1 s exposures for the Hck protein sample, the buffer, and the empty capillary. SAXS profiles were determined by the subtraction of the buffer and the empty capillary scattering from the protein sample scattering. See more details in SI Appendix.

Multidomain Modeling of Hck

Extensive conformational sampling was carried out using a multistate CG model of Hck following methods similar to previous studies (16, 31, 32). Models of inactive and active states of the catalytic domain were built from the crystal structures of Hck (PDB entry: 1Y57) and c-Src (PDB ID: 1Y57), respectively (4, 10), and models of the SH3 and SH2 were taken from the inactive state of the down-regulated Hck (4). The models are built at the residue level, where each residue is represented by its CA atom. For postdata analysis, a clustering and mapping method as described previously was applied to simulation data (17). The resulting configurations were clustered into 25 structural clusters, where a list of 9739binarid pair wise residue distances were used in a K-means clustering calculation. For each cluster, a number of 100 randomly selected configurations was used to compute the theoretical scattering profiles using the Fast-SAXS method (18), yielding 25 profiles (and their standard deviations) log(I(q)/I₀) log(I(q)/I₀) where I₀ = 1,...,25. A second level K-means clustering procedure was performed using a metric based on the similarity between calculated scattering patterns. This second step assigns the 25 structural clusters to N₀ = 9 scattering classes or states (see SI Appendix: Fig. S1). See the SI Appendix for more details.

ACKNOWLEDGMENTS

We thank Markus Seelig and John Kuriyan for providing the pHCK3D and Yohp Duet plasmids and Richard Jones for synthesizing the p2 and p3 peptides. SAXS experiments were carried out at the BioCAT of the Argonne National Laboratory with the help of Dr. Liang Guo. We thank Alper Dagcan for helping with SAXS experiments. This work was supported by the National Institutes of Health via grant CA-093577 from the National Cancer Institute. The computations were supported by Teragrid under project MCA0105108 and by the San Diego Supercomputer Center.

References

Supplementary Information for “Multidomain Assembled States of Hck Tyrosine Kinase in Solution”

Plasmids

The pHCK3D and YopH Duet plasmids were provided by Markus Seeliger and John Kuriyan. pHCK3D is a pET-28a vector (Novagen) modified to yield a tobacco etch virus (TEV) protease cleavable N-terminal hexahistidine tag. It contains a fragment of human HCK spanning the SH3, SH2, and catalytic domains, and the C-terminal tail (residues 85-531, chicken c-Src numbering). YopH Duet is a pCDFDuet-1 vector (Novagen) containing full-length YopH phosphatase from Yersinia. The pHCK3D.YEEI mutant was generated using the QuikChange Site Directed Mutagenesis kit (Stratagene) with pHCK3D as a template and mutagenic primers from Integrated DNA Technologies. This construct contains a high affinity C-terminal tail sequence (E524SQYEEIP531).

Protein expression and purification

Wild type HCK3D or the HCK3D.YEEI mutant were co-expressed in bacteria with YopH phosphatase and purified as described by Seeliger et al. The two plasmids containing the kinase and the phosphatase were co-transformed into Escherichia coli BL21(DE3) cells. 1-liter cultures were grown in Terrific Broth with 50 µg/ml each of kanamycin and streptomycin at 37°C to an OD600 of 1.0, cooled for 1 hour with shaking at 18°C, and induced overnight at 18°C with 1mM IPTG. Cells were harvested by centrifugation at 4000 g for 10 minutes at 4°C. Pellets were resuspended in 30 ml of 20 mM Tris (pH 8.0), 500 mM NaCl, 5% glycerol, 20 mM imidazole (NiA buffer) and lysed by 4 cycles of homogenization at 15,000 psi using an Avestin homogenizer. The lysate was
cleared by centrifugation at 125,000 g for 50 minutes at 4 °C. The supernatant was loaded onto 5 ml NiNTA resin (Qiagen) equilibrated in NiA buffer. After washing with 10 column volumes of NiA buffer, the protein was eluted with 3 column volumes of NiB buffer (NiA buffer plus 500 mM imidazole). The eluted protein was cleaved with 1 mg TEV per 25 mg kinase at 4 °C overnight while dialyzing against 2 liters of 20 mM Tris (pH 8.0), 100 mM NaCl, 5% glycerol, 0.5 mM EDTA, 0.5 mM TCEP in a 12-14 kDa molecular weight cutoff membrane. The dialyzed protein was diluted 1.3 fold with water, spun at 14,000 g for 10 minutes at 4°C, and loaded onto an ion exchange column (HiTrap Q FF, GE Lifescience) equilibrated with 20 mM Tris (pH 8.0), 0.5 mM TCEP (QA buffer). The kinase was eluted with a linear gradient of 0-50% QB buffer (QA buffer plus 1 M NaCl). The peak fractions were analyzed by SDS-PAGE and the fractions containing kinase were pooled. A 1.5 fold molar excess of the inhibitor PP1 (Tocris) was added to prevent aggregation and the pooled fractions were concentrated using an Amicon Ultra-15 centrifugal filter device with a 30-kDa molecular weight cutoff membrane. The concentrated sample was loaded onto a size-exclusion column (Superdex 200 10/300 GL, GE Lifescience) equilibrated with 50 mM Tris (pH 8.0), 100 mM NaCl, and 0.5 mM TCEP. HCK was eluted at the volume expected for monomeric protein. No aggregation was detected in the void volume of the column. Sample concentrations were adjusted to 1 mg/ml for SAXS data collection.

**Peptides**

Peptides were synthesized in house. The SH2 binding peptide (referred to as p2) has the sequence EPQ\[pY\]EEIPIKQ (where pY is phosphotyrosine). The SH3 binding peptide (referred to as p3) has the sequence VSLARRPLPPLP. Stock solutions of p2 (800 µM in Superdex column buffer)
and p3 (8 mM in 50% DMSO) were stored in aliquots at -80°C until needed for SAXS data collection.

**SAXS data collection**

SAXS data were collected at the BioCAT-18ID from the Advanced Photon Source (APS), Argonne, IL. A flow-cell was used to deliver continuous flow through the capillary during data collection to reduce the radiation damage. The temperature of the capillary and sample cell was kept at around 10 °C. Data were collected with a series of 1-second exposures for the Hck protein sample, the buffer and the empty capillary. SAXS profiles were determined by the subtraction of the buffer and the empty capillary scattering from the protein sample scattering.

**Modeling for the multi-domain assembly of Hck**

To explore multiple conformations adopted by the multi-domain Hck complex, extensive sampling and modeling were carried out as follows. For convenience, three segments (Ile$_{82}$-Ser$_{142}$, Glu$_{144}$-Met$_{246}$, and Glu$_{255}$-Thr$_{523}$) in the human Hck are referred to as SH3, SH2, and catalytic-domain, respectively. For inter-domain interactions, the association was calibrated according to dissociation constants ($K_d$) between SH2 and a peptide with sequence of ESQYQQQP, and between SH3 and a peptide with sequence of SSKPQKW. The former peptide resembles the E$_{524}$SQYQQQP$_{531}$ segment in the C-terminal tail of Hck, while the latter resembles the S$_{247}$SKPQKW$_{254}$ sequence in the SH2-connecting polyproline linker region. The reported $K_d$’s are in the µM and tens of µM ranges for SH2-p2 and SH3-p3 2-4, respectively. To implement these inter-domain interactions, umbrella sampling with harmonic biasing potentials was first performed to obtain a potential
of mean force (PMF) of $W(r)$ (where $r$ are inter-residue distances between Asn$_{135}$ in SH3 and Gln$_{251}$ as in the former peptide and between Tyr$_{202}$ in SH2 and Tyr$_{527}$ as in the latter peptide). From $W(r)$, a theoretical $K_d$ was then calculated for the association between each domain and its binding peptide, $K_d = 1/(4\pi \int r^2 e^{-W(r)/k_B T} dr)$, where $k_B$ is the Boltzmann factor and $T$ is the system temperature. The resulting $K_d$’s of 1.2 µM and 27 µM, respectively, were implemented for the intramolecular interactions via the association with SH2 and SH3. To account for competitive binding of external peptides, these intramolecular interactions were also allowed to be on/off in the model. For example, when the interactions of SH2 and the C-tail are turned off, the C-tail is allowed to release from SH2 so that the binding interface of SH2 is available for an external SH2 ligand to bind to the complex. In addition, the SH2-tail interactions are also strengthened to account for its increased binding in a high-affinity mutant with a sequence of E$_{524}$SQYEEIP$_{531}$. In this case, a value of $K_d$ of 0.56 µM between the SH2 and the high-affinity YEEI-containing peptide was used for the SH2 association in the complex.

For the catalytic domain, a multi-state coarse-grained model was applied to account for its internal transition $^5$, where the energy function is built as a combination of two states, $\mathcal{E} = -\log \left( e^{-\beta (E_1 + \delta)} + e^{-\beta E_2} \right) / \beta$ with $\beta = 1/(k_B T)$ $^6$. The parameter of $\delta$ can be viewed as a energy difference between two states, represented by $E_1$ and $E_2$ $^5$. In practice, $\delta$ was used as an advantage for sampling to accelerate the transition from one state to another. Models of inactive and active states of the catalytic domain were taken from the crystal structures of Hck (PDB entry: 1QCF) and c-Src (PDB entry: 1Y57), respectively $^7,8$. The average scattering profile calculated such assembly state is sufficiently different to be “SAXS-distinguishable”, as shown in Fig. 2 Similar to the catalytic domain, a coarse-grained representation for SH3 and SH2 domains was modeled after
their structures \textsuperscript{5,9}. We used the structures of SH3 and SH2 taken from the inactive state of the down-regulated Hck \textsuperscript{7}. The coarse-grained model was built at the residue level where each residue is represented by its C\textalpha atom. It includes bond, angle, and dihedral interactions between adjacent residues, repulsive interactions for residue pairs that are not in contact, and attractive interactions for long-range contacts \textsuperscript{9}.

To sample the landscape of the Hck assembly, a total of 6 sets of simulations with the model and enhanced sampling techniques described above were carried out. Each of them was performed either by turning on/off the intramolecular interactions, by increasing the SH2-tail interactions accounting for mutants, or by adjusting \( \delta \) for sampling (\( \delta = 0 \) when simulations start from the inactive compact or \( \delta = -25 \) kcal/mol when simulations start from a disassembled state). Langevin molecular dynamics were implemented with a friction coefficient of 50 ps\textsuperscript{-1} and a time step of 0.01 ps, and each simulation has 20 \( \mu \)sec. It results in a total of 120 \( \mu \)sec simulation data for further analysis.

For post-data analysis, a clustering and mapping method as described previously was applied to simulation data \textsuperscript{10}. The resulting configurations were clustered into 25 structural clusters, where a list of 89 inter-domain pairwise residue distances were used in a K-means clustering calculation. A resulting transition probability matrix was then constructed to monitor inter-cluster hopping. The use of this transition matrix is two-fold. First, an estimation about the probability of each structural cluster is yielded at a first-order approximation. Obtaining the probability distribution can help solve the problem of local oversampling where certain regimes of the landscape were heavily sampled locally but less favored globally. Second, a connectivity map connecting 25 structural
clusters was constructed from the transition matrix (see Fig. S1). More details on constructing this connectivity map can be found in previous publications.\textsuperscript{5,10}

**SAXS computation and state assignment**

A coarse-grained Fast-SAXS method was subsequently applied to compute theoretical SAXS profiles of all structural clusters.\textsuperscript{11} Fast-SAXS takes advantage of the coarse-grained nature of SAXS data, providing a rapid scattering determination from massive configurations generated from Hck simulations. In the Fast-SAXS calculations, each configuration was soaked in explicit water to account for the contribution from the hydration layer to the scattering. For each cluster, a number of 100 randomly selected configurations was used to compute the averaged scattering profiles (plus their standard deviations). Thus, a total of \(n = 25\) scattering profiles \((\log I_i(q) \pm \delta \log I_i(q)\) where \(i = 1, \cdots, n)\) were achieved.

To measure the scattering similarity between structural clusters, a next level K-means clustering procedure was performed. This scattering clustering was based on pair-wise scattering similarity,\

\[
S_{ij} = \left[ \sum_{q_i} \frac{(\log I_i(q) - \log I_j(q))^2}{(\delta \log I_i(q)^2 + \delta \log I_j(q)^2)} \right]^{1/2} \tag{1}
\]

where \(q_1 = 0.01\) Å and \(q_2 = 0.6\) Å were used and \(\delta \log I_i(q)\) is the uncertainty of scattering intensity \(\log I_i(q)\) of Cluster \(i\). This scattering clustering results in the assignment of the 25 structural clusters into a total of 9 scattering states shown in Fig. S1. Each of the 9 scattering state includes multiple conformers, as shown in Fig. S4 to Fig. S12.


Figure 1: (A) Graph representation of 25 structural clusters and their state assignment into 9 scattering states based on the scattering similarity. (B) Graph representation of the network of Hck assembly states distanced by scattering difference computed by Eq. (1).
Figure 2: The configuration generated by the CG model can be organized into states exhibiting distinguishable SAXS profiles. Shown are the SAXS scattering profile from state 1 (blue solid line) and of state 8 (red dashed line). The SAXS profile $I(q)$ is shown in a logarithmic scale and $q$ is the amplitude of momentum transfer. An efficient Fast-SAXS method was used to compute average scattering profiles of each state \cite{11}. The ensemble-averages (red and blue lines) and uncertainties (black bars) were obtained from 100 randomly selected configurations within each state, respectively. A complete list of $\chi^2$ differences between states is given in Supplementary Table 1.
Table 1: A list of cross-$\chi^2$’s calculated for all the scattering states using Eq. (2) with $q_{\text{min}} = 0.0$ Å and $q_{\text{max}} = 0.15$ Å.

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>113.3</td>
<td>78.4</td>
<td>81.5</td>
<td>164.8</td>
<td>214.9</td>
<td>210.0</td>
<td>356.3</td>
<td>418.9</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>20.6</td>
<td>52.3</td>
<td>128.4</td>
<td>194.2</td>
<td>177.7</td>
<td>299.3</td>
<td>347.9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>38.6</td>
<td>101.2</td>
<td>168.3</td>
<td>145.0</td>
<td>232.4</td>
<td>267.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>48.4</td>
<td>129.2</td>
<td>94.5</td>
<td>158.1</td>
<td>169.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>109.8</td>
<td>60.8</td>
<td>127.4</td>
<td>130.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>68.2</td>
<td>39.4</td>
<td>49.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>53.7</td>
<td>41.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>25.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Theoretical validation of the Bayesian-based Monte Carlo (BMC) simulations. To test the BMC analysis, putative theoretical values of population fractions (blue stems) were used as inputs for the simulations. (a, b) Putative theoretical and BMC-simulated population fractions inferred from the full-set SAXS profiles and (c, d) from the low-q region. The inputs are in blue, while the simulated are in red. (e, f) Similar tests for the input scattering intensity $I(q)$ but with the addition of fluctuations in the amplitude of $\sigma(q)$ and (g, h) in the amplitude of $5\sigma(q)$, where $\sigma(q)$ is the uncertainty of $I(q)$.
Figure 4: Snapshots and an ensemble of structures of state 1, which resemble the crystal structure of the fully assembled state (PDB entry: 1QCF).
Figure 5: Snapshots and the ensemble of structures of state 2 (an SH3-displaced state with a detached C-tail). The observation of this SH3-displaced state is consistent with the previously reported hypothesis.
Figure 6: Snapshots and the ensemble of structures of state 3 (an SH3-displaced state with an associated C-tail).
Figure 7: Snapshots and the ensemble of structures of state 4.
Figure 8: Snapshots and the ensemble of structures of state 5 (disassembled and extended).
Figure 9: Snapshots and the ensemble of structures of state 6 (disassembled and compact).
Figure 10: Snapshots and the ensemble of structures of state 7.
Figure 11: Snapshots and the ensemble of structures of state 8, which resembles the crystal structure of c-Src (PDB entry 1Y57).
Figure 12: Snapshots and the ensemble of structures of state 9.
Figure 13: Histogram of the RMSD of the structures within state 1. A reference structure of the fully assembled structure (PDB entry 1QCF) was used for the RMSD calculations.