Multidomain assembled states of Hck tyrosine kinase in solution

Sichun Yang1,2, Lydia Blachowicz1, Lee Makowski3, and Benoît Roux4,a

1Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637; and 2Biosciences Division, Argonne National Laboratory, Argonne, IL 60439

Edited by John Kuriyan, University of California, Berkeley, CA, and approved July 15, 2010 (received for review April 5, 2010)

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 seeking to optimize the theoretical scattering profiles against experimental SAXS data. The study establishes the concept of basis-set supported SAXS (BSS-SAXS) reconstruction combining computational and experimental techniques. Here, BSS-SAXS reconstruction is used to reveal the structural organization of Hck in solution and the different shifts in the equilibrium population of assembly states upon the binding of different signaling peptides.

Tyrosine kinases of the Src-family are large multidomain allosteric enzymes implicated in the signaling pathways regulating cell growth and proliferation (1). The key role that Src kinases play in the onset of many human diseases, particularly cancer, makes them important targets for therapeutic intervention (2).

All Src kinases share a common structural organization comprising the SH3 and SH2 binding domains followed by a highly conserved catalytic domain connected by flexible linkers (1). Crystal structures have offered a detailed view of the down-regulated forms of Hck and c-Src (3, 4), characterized by a compact assembled form stabilized by auto-inhibitory intramolecular interactions between the N and C terminus of the catalytic domain with the SH3 and SH2 modules, respectively. The SH2 and SH3 modules of the Src-family tyrosine kinases play dual roles: both are involved in the auto-inhibitory intramolecular interactions down-regulating kinase activity, but can also serve as possible binding receptors for cellular signals leading to kinase activation.

In a broadly accepted paradigm, Src inactivation/activation is pictured in terms of a simple two-state process involving an assembled (inactive) and a disassembled (active) conformation, i.e., once any of the intramolecular interaction is released, the kinase domain switches to a disassembled catalytically active state targeting downstream cellular substrates (5–8). Crystal structures (3, 4) and mutational studies complemented by molecular dynamics (MD) simulations (9) are consistent with this view. However, one crystal structure of c-Src in a partially activated state in which the SH2-SH3 modules are reassembled in a different orientation with respect to the catalytic domain (10), suggests that the situation could be considerably more complex. In this regard, it seems likely that different signals could promote different assembly states, leading to differently configured activated kinases. While it is understood that Src increases its activity in response to molecular signals that interfere with the interactions stabilizing the assembled complex, little is known about the spatial organization of the up-regulated active form of Src. A quantitative characterization of the structural organization of the multidomain Src in solution under different conditions is critical for a better understanding of the molecular basis of the allosteric response.

Probing the spatial organization of a large and flexible multidomain macromolecular complex such as Src is challenging. The protein is expected to display substantial conformational dynamics, giving rise to a large number of possible conformations separated by small energy differences. Capturing any of those in a crystallographic state would be inherently difficult, and ideally one would like to observe the protein conformation in solution, where it is not affected by lattice packing. In principle, small-angle X-ray solution scattering (SAXS) is an experimental technique that can be used for mapping the three-dimensional organization of multidomain proteins in solution (11–15). Different conformations give rise to distinct scattering patterns containing information on the overall architecture and domain-domain separation within the complex. However, it is not straightforward to determine the three-dimensional shape directly from the low-resolution information provided by the SAXS data. More importantly, there are fundamental limitations associated with the naive notion of a unique molecular shape in the case of flexible multidomain complexes, which may display considerable conformational dynamics.

To address those issues, we present an integrative approach, the basis-set supported SAXS (BSS-SAXS) reconstruction, that aims at combining SAXS data with a simplified coarse-grained (CG) model to characterize the assembly states of the flexible multidomain Hck in solution. Simulations of the CG model are used to extensively explore and sample the accessible conformational space of the multidomain complex. The configurations are clustered into a small number of distinct putative assembly states, which are then used as a basis-set to analyze SAXS data. Comparison of the calculated SAXS patterns from all assembly states with experimental data makes it possible to determine the population fraction of each state of Hck under various conditions. 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.

Results

Classes of Hck Assembly States. The simplified CG model of Hck, though extremely simplified, incorporates the main generic features of the protein: (i) the native three-dimensional conformations of the SH3, SH2, and catalytic domains from the X-ray structure of human c-Src (16) and (ii) the basis-set supported SAXS structure of the phospho-activated form of c-Src (13). The authors declare no conflict of interest.

Author contributions: S.Y. and B.R. designed research; S.Y. performed research; S.Y. and L.B. contributed new reagents/analytic tools; S.Y. 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.

†Present address: Center for Proteomics and Bioinformatics and Department of Pharmacology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004569107/-/DCSupplemental.

Published September 7, 2010.

www.pnas.org/cgi/doi/10.1073/pnas.1004569107
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 polyproline SH2-binding “linker” with the SH3 domain. As described in SI Appendix, both interactions were incorporated in the CG model, and adjusted to match the data from experiments for the SH2-binding Ctail peptide, and the SH3-binding polyproline 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 detailed 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 \( N_i = 9 \) states, weighted by the fractional population \( \{P_i\} \equiv P_1, \ldots, P_N \), of state \( i \),

\[
I_{\text{calc}}(q) = \sum_{i=1}^{N} \frac{P_I}{P_i}(q),
\]

where \( P_I(q) \) 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 \( N_i \), 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,

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

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 \( \sqrt{\log I_{\text{exp}}(q)} \). Assuming independent experimental errors and a uniform prior, the probability distribution \( \exp(-\chi^2/2) \) 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 \( N_i \) states, along with their uncertainties extracted the fluctuations.

Tests with the BSS-SAXS method indicate that the scattering pattern in the low-\( q \) regions, up to about 0.15 Å\(^{-1}\) contains most of the information that controls the structural differences in Hck (see SI Appendix: Fig. S3A-C). Including higher-\( q \) SAXS data has little effect on the average values of \( P_i \), although it reduces the uncertainty of the \( P_i \) (see SI Appendix: Fig. S3B, D). Additional tests with the BSS-SAXS/BMC simulations show that the procedure yields the correct population \( P_i \) (see SI Appendix: Fig. S3E, F), although large increases in the estimated SAXS uncertainty can lead to inaccuracies (see SI Appendix: Fig. S3G, H). The present approach has some similarities with other methods, such as EOM (19) and BILBOMD (20), which were also designed to go beyond standard strategies based on a single conformation to fit SAXS data. For example, all three methods primarily rely on a large set of random conformers. Both EOM and BILBOMD use a Monte Carlo genetic algorithm to select a small number of specific conformers out of the large set to best-fit the SAXS data, which are attributed equal statistical weight, whereas BSS-SAXS uses a BMC algorithm to determine the fractional population \( P_i \) for the states in a scattering basis-set. Overall, the tests show that the BSS-SAXS analysis can provide quantitative estimates of the fractional population \( P_i \) of Hck in solution.

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...
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 SASX data based on the variations of \( R_e \) 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 SASX 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 SASX 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 SASX patterns \( I(q) \) (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 SASX data for the wild-type Hck is shown in Fig. 2A. The analysis indicates that the enzyme is predominantly in the assembled state 1, with a probability of 82%. As shown in the figure (left), the calculated scattering profile with the combination of states determined from the BMC analysis (red curve) is in excellent agreement with the experimental profile with the combination of states determined from the BMC analysis (black curve). Although the assembly state 1 is similar to the X-ray structure of Hck (3, 4), the ensemble of configurations includes dynamical fluctuations with rmsd on the order of 2–3 Å (see SI Appendix: Fig. S4). The scattering pattern calculated from the static X-ray structure alone does not fit the SASX data accurately. In fact, the assembly state 1 (with dynamical fluctuations) also yields a poor score (Fig. 2A, right). The contribution from assembly states 6 and 8 is critical and improves the goodness of fit. In a previous SASX study of c-Src, it was concluded that about 83% of the enzymes were in a conformation similar to the assembled form (25). In those experiments, Tyr\(_{527}\) in the C terminus of the catalytic domain was phosphorylated by the kinase Csk, which is required for kinase inactivation. It is particularly noteworthy that the assembled form is predominant, even though Tyr\(_{527}\) is unphosphorylated in the present experiments. The interaction of the phosphorylated tyrosine with the SH2 module is critical to stabilize the assembled form of the kinase (3, 4). It is well known that dephosphorylation of Tyr\(_{527}\) leads to kinase activation and truncation of the C terminus leads to a constitutively active kinase. Yet, the present results indicate that the assembled conformation is also predominant in solution (84%), even when the tyrosine in the Ctail is unphosphorylated. Thus, despite the expectation that Hck with an unphosphorylated Tyr\(_{527}\) in the Ctail might adopt a multiplicity of disassembled states, the assembled conformation typical of the inactive kinase is predominant. The present observations challenge the notion that dephosphorylation of the Ctail immediately triggers a large conformational change toward a disassembled state leading to kinase activation.

Comparison of the results obtained on the wild-type and mutant kinase sheds an interesting light on the competition between the binding of the Ctail and an external p2 peptide, corresponding intramolecular and bimolecular association processes, respectively. The experimental SASX data from the high-affinity Ctail mutant Hck-YEEI is shown in Fig. 2B. Here 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 \( R_e \) 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-YEEI 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-YEEI mutant. This destabilization is also reflected in the \( R_e \) of 29.3 Å and 28.1 Å, for the wild-type and Hck-YEEI 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-YEEI mutant and, to a lesser extent, wild-type Hck compete against the binding of a very high-affinity peptide and protect the assembled 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 governed by \( [D]/[A]=p \), and \( [DL]/[D]=L/K_d \), where \( p \) is the natural propensity of occurrence of the D state, and \( K_d \) is the dissociation constant, respectively. The probability of the assembled state is then equal to, \( P_A = (1+p) L / K_d \). Because \( p \) is smaller for the high-affinity Ctail mutant Hck- YEEI than for the wild-type Hck, the intramolecular interaction is able to override the strong bimolecular association (\( p \sim 10^{-6} \)), 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 SASX data for the Hck-YEEI 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 basis 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 YopH phosphatase and purified as described by Seeliger et al. (30). The SH2 binding peptide (referred to as p2; sequence EPO[pY]EIEI) and the SH3 binding peptide (referred to as p3; sequence VSLARRPLPPLP) were synthesized in house. SAXS data were collected at the BioCAT-18ID 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 configurational 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: 1QCQ) and c-Src (PDB entry: 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 backbone.

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 89 interdomain pair wise residue distances were used in a K-means clustering algorithm. For each cluster, a number of 100 randomly selected configurations was used to compute the theoretical scattering profiles using the Fast-SAXS method (18), which was fitted to the experimental profiles (and their standard deviations) log|I(q)|=log|I(q0)|+q2δα(q2−q20)+⋯, where J = 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 into Nclustering classes or states (see SI Appendix: Fig. S1). See the SI Appendix for more details.

ACKNOWLEDGMENTS. We thank Markus Seeliger and John Kurian for providing the pHCK3D and YopH Duet plasmids and Richard Jones for synthesizing 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 MCA010518 and by the San Diego Supercomputer Center.