Replica exchange simulations of transient encounter complexes in protein–protein association

Young C. Kim, Chun Tang, G. Marius Clore*, and Gerhard Hummer*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Edited by Axel T. Brunger, Stanford University, Stanford, CA, and approved June 24, 2008 (received for review March 11, 2008)

Recent paramagnetic relaxation enhancement (PRE) studies on several weakly interacting protein complexes have unequivocally demonstrated the existence of transient encounter complexes. Here, we present a computational method to study protein–protein binding by creating equilibrium ensembles that include both specific and nonspecific protein complexes. In a joint analysis of simulation and experiment we explore the physical nature and underlying physicochemical characteristics of encounter complexes involving three protein–protein interactions of the bacterial phosphotransferase system. Replica exchange Monte Carlo simulations using a coarse-grained energy function recover the structures of the specific complexes and produce binding affinities in good agreement with experiment. Together with the specific complex, a relatively small number of distinct nonspecific complexes largely accounts for the measured PRE data. The combined relative population of the latter is less than ~10%. The binding interfaces of the specific and nonspecific complexes differ primarily in size but exhibit similar amino acid compositions. We find that the overall funnel-shaped energy landscape of complex formation is dominated by the specific complex, a small number of structured nonspecific complexes, and a diffuse cloud of loosely bound complexes connecting the specific and nonspecific binding sites with each other and the unbound state. Nonspecific complexes may not only accelerate the binding kinetics by enhancing the rate of success of random diffusional encounters but also play a role in protein function as alternative binding modes.

protein binding | transient nonspecific encounter complexes | paramagnetic relaxation enhancement | bacterial phosphotransferase system

Experimental and theoretical studies have provided evidence that transient nonspecific encounter complexes play an important role in protein binding and function (1–17). The formation of weakly bound nonspecific complexes, often dominated by long-range electrostatic interactions, enhances the on-rate of binding by increasing the interaction cross-section and reducing the conformational space to be searched on the path to the specific complex. From the thermodynamic standpoint, weak nonspecific complexes contribute to the overall affinity by bolstering contacts between proteins or domains (14, 17, 18). Despite the importance of encounter complexes, little is known about their relative populations or structures, largely because of their transient nature.

Recent work has shown that paramagnetic relaxation enhancement (PRE) provides a very sensitive tool for detecting the presence of low population transient species in solution. PRE measurements have unequivocally demonstrated the existence of transient encounter complexes in protein–DNA (19, 20) and protein–protein (4–8, 21) interactions. These experiments rely on three key observations: (*i*) the observed intermolecular PRE relaxation rates in the fast-exchange regime are population averages of all complexes present in solution; (*ii*) the magnitude of the PRE is proportional to $\langle r^{-6} \rangle$, where r is the distance between a proton and a covalently attached paramagnetic label; and (*iii*) PRE effects are strong owing to the large magnetic moment of an unpaired electron.

Here, we present a computational method, based on rigid-domain motion and residue-level potential energies, for simulating transient protein encounter complexes. In a joint analysis, we combine simulation and experiment to characterize the relative populations and structures of encounter species. As test cases, we consider three protein–protein complexes involved in the bacterial phosphotransferase system, namely the complexes of the histidine phosphocarrier protein HPr with the N-terminal domain of enzyme I (EIN), $IIA^{Mannitol}$ (IIA^{Mtl}), and $IIA^{Mannose}$ (IIA^{Man}). We make use of replica exchange Monte Carlo (REMC) simulations of the complexes using a coarse-grained protein model with a transferable potential energy function (18, 22) that combines long-range electrostatics with short-range statistical contact potentials (23). This model and energy function are suited for studies of weakly binding protein complexes ($K_D > 1$ μM) and have been tested against structure and affinity measurements (18, 22).

The paper is outlined as follows. We first assess the simulation model by comparing the calculated complex structures and binding affinities to experiment. We then show that the nonspecific complexes contained in the simulation ensembles can qualitatively account for key features in the measured PRE profiles. To estimate the relative populations of the encounter complexes, we cluster the simulation structures and reweight the clusters to match the measured PRE profiles. We then characterize the resulting ensemble of encounter complexes with regard to their population, structure, and interfacial characteristics. In conclusion, we summarize the key findings and discuss their implications for the functional role of encounter complexes.

Results and Discussion

Structure and Binding Affinity. To assess the quality of the simulation model, we first compare the calculated binding affinities and complex structures to experiment. To obtain equilibrium ensembles of bound and unbound complexes, we perform long REMC simulations of unlabeled HPr in cubic, periodically replicated boxes together with one of its binding partners (EIN, IIA^{Man}, or IIA^{Mtl}). For each of the three complexes, equilibrium ensembles are collected for box sizes ranging from 200 to 1,500 Å, corresponding to protein concentrations of 0.5–208 μM. Binding affinities from structural and calorimetric approaches are found to agree (Fig. 1*A*). In the calorimetric approach, the peak in the excess heat capacity as a function of protein concentration yields the binding affinity (Fig. 1*A Inset*). In the structural approach, the fraction of HPr bound to its counterpart is fitted to a titration curve as a function of protein concentration [see Fig. 1 for the EIN-HPr and IIA^{Man}-HPr complexes and

Author contributions: Y.C.K., C.T., G.M.C., and G.H. designed research; Y.C.K. performed research; Y.C.K., C.T., G.M.C., and G.H. analyzed data; and Y.C.K., C.T., G.M.C., and G.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

*To whom correspondence may be addressed. E-mail: gerhard.hummer@nih.gov or mariusc@mail.nih.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802460105/DCSupplemental.

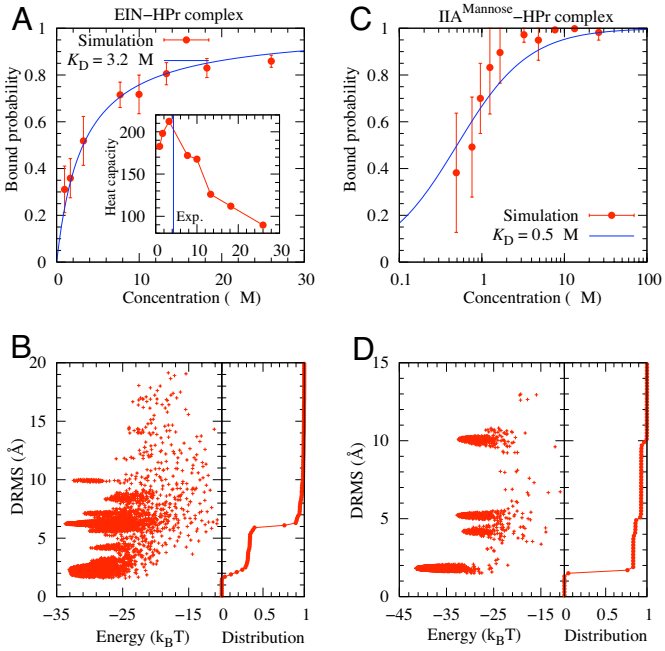

Fig. 1. Complex affinity and structure. (*A*) Fraction of bound states as a function of protein concentration for the EIN-HPr complex derived from the REMC simulations. The solid curve is a titration fit with $K_D = 3.2$ μM. (*Inset*) The excess heat capacity as a function of protein concentration (red symbols) and the experimental K_D (vertical line). (*B*) Scatter plot of distance root mean square (DRMS) versus energy for the simulated EIN-HPr complexes and the corresponding DRMS cumulative distribution. (*C*) Fraction of bound states as a function of protein concentration for the IIA^{Man}-HPr complex together with a titration fit ($K_D = 0.5$ μM). (*D*) Scatter plot of DRMS versus energy for the simulated IIA^{Man}-HPr complexes and the corresponding cumulative distribution.

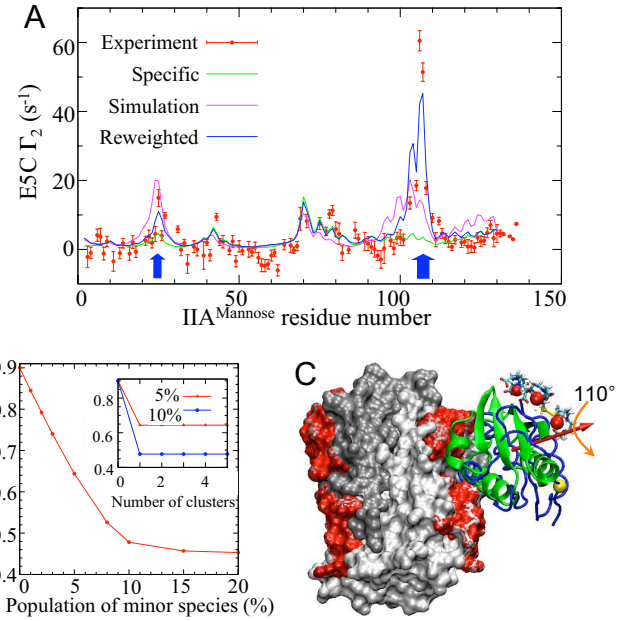

Fig. 2. IIA^{Man} dimer-HPr(E5C) complex. (*A*) Intermolecular transverse ^1H_N-Γ_2 PRE rates from experiment (red circles) (4), simulation (purple), reweighted simulation (blue), and the stereospecific complex alone (green). The blue arrows indicate regions where discrepancies between the observed and stereospecific PRE rates are large. (*B*) PRE *Q*-factor (27) as a function of the population of nonspecific complexes, p_{minor}. (*Inset*) The PRE *Q*-factor as a function of the number of clusters at $p_{minor} = 5$ and 10%. (*C*) Structures of the stereospecific complex (green) and transient encounter complex (blue), with the symmetry-related complexes not shown. The IIA^{Man} dimer is colored in light and dark gray with residues exhibiting PRE rates $>10\,s^{-1}$ in red. The three Mn^{2+} conformers (conjugated to the E5C site) of the specific complex are represented as red spheres, and the E5C site of the nonspecific complex is shown in yellow. The red arrow indicates the axis of the 110° rotation connecting the specific and nonspecific HPr complex structures.

supporting information (SI) *Appendix* and Fig. S1 for IIA^{Mtl}-HPr]. The two proteins are assumed to be bound if the distance between one or more pairs of residues is less than a cutoff distance of 8 Å. The estimated binding affinity for the EIN-HPr complex, $K_D \simeq 3.2$ μM, is in excellent agreement with the experimental value of 4.3 μM (5). For the IIA^{Man}-HPr complex (24), the estimated $K_D \simeq 0.5$ μM is lower than the observed one (~30 μM). Note, however, that a 60-fold difference in K_D amounts to an error of k_BT ln 60 ≃ 2.5 kcal/mol, which may not be resolved within a coarse-grained representation of protein–protein interactions (22) (k_B is Boltzmann's constant and T is the absolute temperature). For the IIA^{Mtl}-HPr complex (25), the calculated and measured binding affinities agree well, $K_D \simeq 15$ and ~50 μM, respectively (*SI Appendix*).

To compare the simulated structures with those from NMR experiments (24–26), we make use of the distance root mean square (DRMS) deviation metric (see *Materials and Methods*). As shown in Fig. 1, the simulations yield bound complex structures close to the stereospecific complexes (DRMS ≈2 Å). Among the bound EIN-HPr and IIA^{Man}-HPr complexes, ~35% and ~90% are native-like with DRMS values <5 Å, respectively. Moreover, in the EIN-HPr complex, HPr binds preferentially to the native binding interface of EIN in >90% of the bound complexes, correctly identifying the binding interfaces for both EIN and HPr with DRMS values <7 Å. The remaining complexes (≲10%) exhibit binding interfaces different from the stereospecific complex. For the IIA^{Mtl}-HPr complex, ≈20% of the simulation structures have DRMS values <5 Å (Fig. S1). The majority of the remaining structures share the binding interfaces with the stereospecific complex but differ in the relative protein orientation.

PRE Profiles from Simulation and Experiment. A comparison of measured and calculated transverse PRE rates allows us to assess the applicability of the simulation procedure to study the structure and population of transient encounter complexes. Intermolecular transverse PRE rates are calculated from the ensemble average of $\langle r^{-6} \rangle$, where r is the distance between the amide hydrogen atoms and the Mn^{2+} paramagnetic label. To account for flexibility, we assume a Gaussian distribution of the Mn^{2+} label around the center of a three-conformer representation of the EDTA-Mn^{2+} group conjugated to an engineered solvent-exposed cysteine residue (4, 27) (see *SI Appendix* and Table S1). Because the simulations were performed without Mn^{2+} labels, simulated structures with overlap between the HPr paramagnetic labels and the protein-binding partner were eliminated from the respective ensembles. A cutoff distance of 8 Å between the center of the Mn^{2+} sites and any residue of the binding partner eliminates <1% and ≲10% of the bound structures for the IIA^{Man}-HPr and EIN-HPr complexes, respectively.

Fig. 2 compares the measured intermolecular PRE profiles for the IIA^{Man}-HPr complex (4) with those calculated from the stereospecific complex alone, and from the ensemble average over the REMC simulations. In the measurements, HPr was labeled with EDTA-Mn^{2+} at residue 5 (E5C). The stereospecific complex does not account for the observed PRE peaks near residues 20–30 and 100–110, as indicated by blue arrows. Remarkably, the PRE rates averaged over the simulation produce

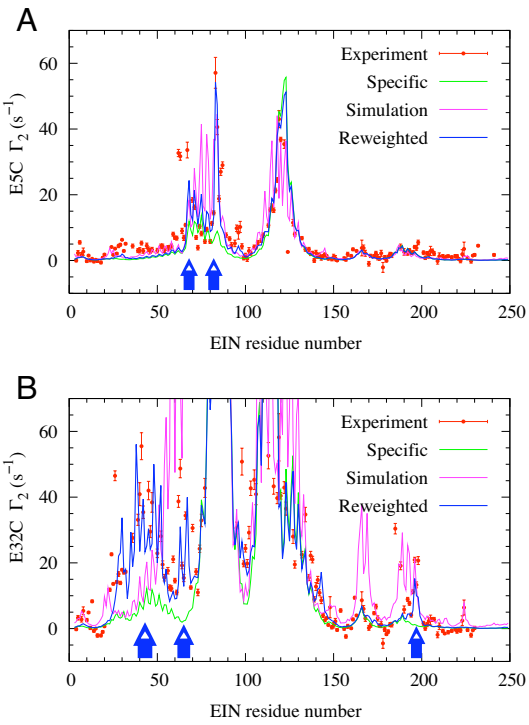

Fig. 3. Intermolecular ^1H_N-Γ_2 PRE rates for the EIN-HPr complex. Results are shown for EIN-HPr(E5C-EDTA-Mn^{2+}) (*A*) and EIN-HPr(E32C-EDTA-Mn^{2+}) (*B*) from experiment (red circles; PRE measurements at 100 mM NaCl were carried out as described in ref. 4), REMC simulations (purple), and reweighted simulations (blue). Blue arrows indicate regions where discrepancies between the observed PRE rates and those calculated for the stereospecific complex (green) are large.

PRE peaks in both those regions, while maintaining the peaks associated with the specific structure.

Averaging of the PRE rates over the simulation ensemble produces a similar improvement for the EIN-HPr(E5C-EDTA-Mn^{2+}) complex when compared with the stereospecific complex (Fig. 3*A*). The simulated PRE rates produce large amplitudes near residues 70–85, consistent with experiment, while otherwise maintaining the profile of the specific complex. For the EIN-HPr(E32C-EDTA-Mn^{2+}) complex, the simulation results are less clear, with improvements over the specific complex near residues 50–60 and 190–200 but significant deviations in other regions (Fig. 3*B*).

We conclude from the analysis of Fig. 2*A* and Fig. 3 that the simulations of the coarse-grained protein model can, at least qualitatively, capture both specific and nonspecific protein complexes, as assessed by the PRE profiles. In particular, the simulations add PRE signals where experimental PRE rates deviate most strongly from PRE rates calculated for the stereospecific complex. However, the agreement between the simulation-averaged PRE rates and experiment is not quantitative. Although the simulated PRE rates for the IIA^{Man}-HPr(E5C-EDTA-Mn^{2+}) complex follow the trend of the experimental data (see Fig. 2*A*), differences between the simulated and experimental PRE rates are apparent near residues 95 and 110. As a consequence, the PRE *Q*-factor (see refs. 4 and 27 and *Materials and Methods*) of the simulations is minimally improved over the stereospecific complex (0.83 vs. 0.9). For the EIN-HPr(E5C-EDTA-Mn^{2+}) complex, the PRE *Q*-factor improves from 0.67 for the specific complex to 0.5, with simulated structures overestimating PRE rates arising from the E5C-EDTA-Mn^{2+} label near residue 70. Despite some qualitative improvement for EIN-HPr(E32C-EDTA-Mn^{2+}), the simulation strongly overestimates PRE rates in the region of residues 70–110, which leads to a large PRE *Q*-factor of 12.6, compared with $Q = 0.7$ for the specific complex. Deviations may in part be caused by approximations in our procedure (in particular, not including the Mn^{2+} labels in the simulations). In addition, the underlying coarse-grained energy function ignores atomic detail (13, 28), and the resulting small k_BT-scale errors can significantly affect the relative populations.

Populations of Transient Encounter Complexes. To obtain relative populations of the specific complex and the different nonspecific encounter complexes, we determined relative weights by fitting the calculated PRE rates to experiment. First, each simulated complex structure was refined via Monte Carlo (MC) energy minimization to locate the closest local minimum on the energy landscape. The energy-minimized structures were clustered by using the DRMS metric, and the average PRE rates were calculated for each cluster. The cluster populations were then reweighted to fit the experimental PRE data by minimizing the PRE *Q*-factor (see *Materials and Methods*). To account for the known structure of the stereospecific complex, the calculated PRE rates from the specific complex were also included.

For IIA^{Man}-HPr(E5C-EDTA-Mn^{2+}), adding a single nonspecific complex with a population p_{minor} of ≈10% greatly improves the PRE *Q*-factor (Fig. 2*B Inset*). Including additional clusters did not significantly improve the agreement. The structures of the specific and nonspecific complexes (without their symmetry-related equivalents) are presented in Fig. 2*C*. Remarkably, we find that the nonspecific complex sampled by the E5C-EDTA-Mn^{2+} paramagnetic label shares its binding interface with the specific complex, differing only in the orientation of HPr.

To ascertain the population of the nonspecific complex, the PRE *Q*-factor was minimized at different relative populations p_{minor} between 0% and 20%. As seen in Fig. 2*B*, a population of $p_{minor} = 10\%$ can account semiquantitatively for the experimental data, with a PRE *Q*-factor of ≈0.45, compared with 0.9 for the specific complex alone. The resulting PRE profile is shown in Fig. 2*A*. The two nonspecific PRE peaks indicated by the blue arrows are well reproduced by the single nonspecific complex shown in Fig. 2*C*. Small deviations between measured and calculated PRE rates may indicate the presence of additional nonspecific complexes with very low affinity.

Using the same procedure, we also refined and clustered the simulated complexes of EIN-HPr. Here, we obtained ~80 clusters. This large number of clusters (compared with ~10 for the IIA^{Man}-HPr complex) is evident in the broader distribution of DRMS values of Fig. 1*B*. By using the PRE rates calculated from these clusters and from the stereospecific complex, the population of each cluster is optimized at a given p_{minor} by globally minimizing PRE *Q*-factors for HPr paramagnetically labeled at the E5C and E32C sites. As shown in Fig. S2, a ≈5–10% nonspecific population is needed to lower the *Q*-factor to ≈0.45 for both the E5C and E32C labels. Only a relatively small number of eight nonspecific complexes are needed to account globally for the PRE data (with one set of three clusters contributing primarily to the E5C profile, and the other five to the E32C profile). The PRE rates calculated with $p_{minor} = 10\%$ are presented in Fig. 3 and show good agreement with experiment. In particular, all of the PRE peaks indicated by blue arrows are well reproduced by the reweighted PRE rates. The EIN-HPr(E25C-EDTA-Mn^{2+}) data (4) were not included in the analysis because the stereospecific complex accounts for the bulk of the measured PRE rates, and of the eight nonspecific clusters, some result in steric clashes of the label whereas the others do not produce any appreciable PRE.

For the IIA^{Mtl}-HPr complex, the reweighting procedure shows that two nonspecific complexes together with the specific com-

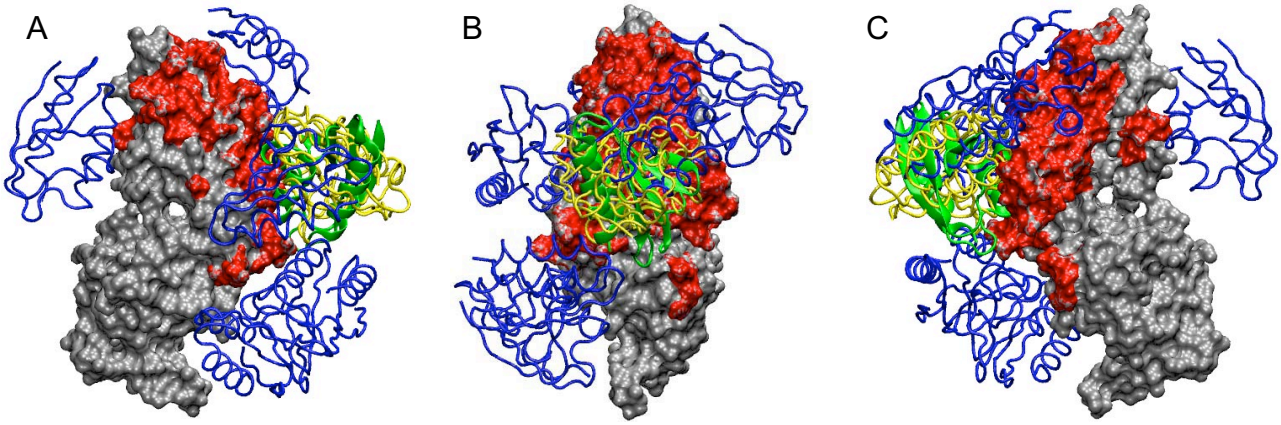

Fig. 4. Structures of the eight dominant nonspecific EIN-HPr complexes derived from the simulations. EIN is colored in gray and the specific structure of HPr is colored in green. EIN residues with transverse PRE rates of $>10\ s^{-1}$ for the E5C label and $>20\ s^{-1}$ for the E32C label are colored in red. Structures colored in yellow and blue are the nonspecific complexes that contribute mainly to PRE profiles observed for the E5C and E32C sites, respectively. Three different views are shown in *A*–*C*.

plex can largely account for the measured PRE data (Fig. S3). The nonspecific complexes have a combined relative population of ~5% and are found to share the IIA^{Mtl} binding interface with the specific complex.

Structural Characteristics of Transient Encounter Complexes. Remarkably, several of the nonspecific complexes occupy the same binding interface as the specific complex, albeit with different relative orientations of the proteins. As discussed above, and shown in Fig. 2*C*, the specific and nonspecific IIA^{Man}-HPr complexes maintain the contact interfaces on both binding partners but have HPr rotated by ≈110°. Similarly, the three nonspecific complexes accounting primarily for the E5C PRE profile of the EIN-HPr complex share the binding interface with the specific complex, as shown in Fig. 4. However, the remaining five complexes deviate significantly from the specific structure.

The nonspecific binding sites also have chemical compositions similar to the specific interface, with the ratio of polar and apolar buried surface being ≈1 to 2 in all cases (Table 1). The main difference is that the overall amount of buried solvent-accessible surface area of the nonspecific complexes is ≈65% of that of the specific complex. This result suggests that the most populated encounter complexes of the bacterial phosphotransferase system are also stabilized by hydrophobic interactions, in addition to electrostatics (5). Earlier studies on different systems have found transient encounter complexes dominated by electrostatic interactions (1, 2, 15, 16), but hydrophobic stabilization has been reported, e.g., for transcription factor binding (9).

Conclusions

By treating folded protein domains as rigid bodies and using a coarse-grained, residue-based energy function, we were able to create equilibrium ensembles of protein complexes that contain both the specific complex and multiple nonspecific transient encounter complexes. In replica exchange Monte Carlo simulations, we recovered the structures of the specific complexes of the histidine phosphocarrier protein HPr with its binding partners EIN, IIA^{Mtl}, and IIA^{Man} at DRMS values <5 Å with relative populations of ≈35%, 25%, and 90%, respectively. Moreover, the calculated binding affinities were found to be in good agreement with experiment, with binding free energy differences of 0.2, 0.7, and 2.5 kcal/mol, respectively. For the IIA^{Man}-HPr complex, the PRE profile measured by NMR is reproduced by the simulations essentially without adjustments. For the other two complexes, we reweighted the relative populations of the different encounter complexes to match the experimental PRE data, allowing us to identify the dominant encounter complex structures. Although the coarse-grained model was able to produce a significant population of native-like structures with DRMS <3 Å for all three cases, the precise structures and relative populations of such structures (as compared with other binding modes) are beyond the scope of the model. In particular, our coarse-grained model neglects details of the atomic interactions, as evident from the relatively higher energy of the native-like EIN-HPr complex compared with some nonnative ones. In such cases, the knowledge of the native structures is helpful, if not necessary, to reweight the clusters of encounter complexes.

The picture that emerges from our analysis of the NMR PRE data is that the ensemble of nonspecific protein encounter complexes can be grouped into two major classes in an overall funnel-like (29) energy landscape of binding (Fig. 5). The

Table 1. Buried solvent-accessible surface area (in Å²) of the stereospecific and top-ranked nonspecific complexes for the IIA^{Man}-HPr and EIN-HPr complexes

	IIA^{Man}-HPr			EIN-HPr		
Complex	Total	Polar	Apolar	Total	Polar	Apolar
Specific	1,493	504	989	1,879	629	1,250
Nonspecific	938	239	699	1,188	396	792

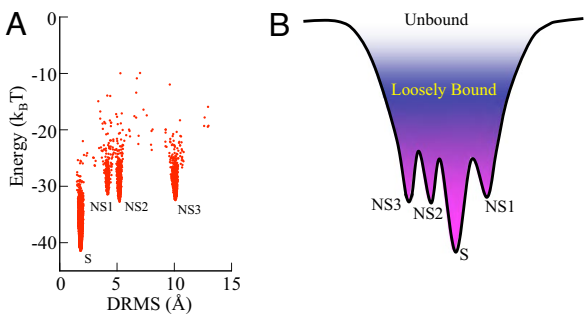

Fig. 5. Binding energy landscape. (*A*) Scatter plot of energy versus DRMS of the IIA^{Man}-HPr complex. (*B*) Schematic funnel representation of the landscape with the specific complex (S) and nonspecific complexes (NS1, NS2, and NS3) connected by a diffuse cloud of loosely bound structures.

dominant group consists of well structured complexes with binding interfaces similar to those of the specific complexes (and, in some cases, overlapping with them). For the weakly binding proteins studied here, with $K_D \approx 1\ \mu$M, the combined population of structured nonspecific complexes is estimated at ~10%, but individually the relative populations of the alternative binding modes are small. The second group consists of loosely bound complexes. In the simulations, these less structured complexes are, effectively, en route between the various specific and nonspecific binding sites, and the unbound state.

The binding interfaces of the nonspecific complexes found here agree well with those identified in the original study (4). However, from simulations using physical interaction models we find here tighter and more structured nonspecific complexes, with buried surface areas ~65% of that of the specific complex. In contrast, the simulated annealing refinement against the PRE data of ref. 4, designed to delineate and visualize the distribution of the nonspecific complexes, produced weaker complexes with a wide range of buried surface areas that were on average one order of magnitude smaller than that of the specific complex. Such loosely bound nonspecific complexes are seen in the simulations, with binding energies above both the specific and nonspecific structured complexes (Fig. 5) and may account for the remaining small deviations between calculated and measured PREs. Indeed, in the original study (4), 2-fold lower *Q*-factors were obtained for a less-restricted ensemble of encounter complexes. Despite the qualitative agreement between the two ensembles of encounter complexes, the observed differences suggest that a small set of one-dimensional PRE profiles is not fully sufficient to define a unique ensemble of three-dimensional encounter complexes. To fully delineate the ensemble of encounter complexes, PRE profiles with many labels are needed.

Evidently, the presence of weak, alternative binding sites on the protein surfaces, connected by diffuse clouds in the ensemble of complex structures, should affect the on-rate of binding. The alternative interaction possibilities greatly enhance the chance that an initial diffusive encounter leads to successful binding, whereas the binding strength of the nonspecific sites is weak enough to ensure that the complex partners are not trapped on their diffusive route to the specific site.

These results raise some interesting biological questions: First, is the possible acceleration of the on-rate through the formation of nonspecific complexes an evolutionary factor, i.e., have the sequences evolved to maintain a certain amount of nonspecificity in the interactions? Second, are the distinct nonspecific binding modes of possible functional relevance? And finally, are those alternative sites accidental, or are they remnants of earlier binding modes in evolutionary history that may still be populated in distant relatives? Answers to these questions may emerge as more protein complexes are analyzed experimentally and careful sequence analyses are performed to shed light on the delicate balance between specific and nonspecific interactions.

Mutations at the native binding interface may shift the population balance between the different binding modes. In particular, it may be possible to depopulate the specific site and increase the population in nonspecific sites to enable structure determination by x-ray crystallography or conventional NMR spectroscopy. Indeed, such an approach has already led to a series of different structures of the complex between cytochrome *c* peroxidase and cytochrome *c* (30), consistent with binding modes seen in recent PRE measurements (7). Interestingly, the different binding modes share the binding interface on cytochrome *c* peroxidase but have cytochrome *c* in different orientations. Similar behavior was seen here for several nonspecific complexes, and sharing of interfaces was also reported for proteins that bind multiple targets (31). In the case of the interaction of IIA^{Man} with its downstream partner $IIB^{Mannose}$, a mixture of productive (i.e., phosphoryl-transfer-competent) and nonproductive complexes have been observed experimentally by NMR, with the alternative binding mode of the nonproductive complex possibly relevant for subsequent steps in the phosphoryl-transfer cascade (32).

Materials and Methods

Coarse-Grained Model. The protein complexes were simulated by adapting the coarse-grained model of ref. 22 with the proteins treated as rigid bodies. The residue-specific pair interaction potentials $\varphi_{ij}(r)$ (where r is the distance between the C^α atoms) combine long-range electrostatics with short-range interactions, $\varphi_{ij}(r) = u_{ij}(r) + u_{ij}^{el}(r)$. The short-range interactions are represented by a 12-10-6 potential form (33),

$$u_{ij}(r) = |\varepsilon_{ij}|[13(\sigma_{ij}/r)^{12} - 18(\sigma_{ij}/r)^{10} + 4(\sigma_{ij}/r)^6]\, f(r), \quad [1]$$

where ε_{ij} and σ_{ij} are the residue-dependent interaction strengths and radii, and $f(r) = 1/[1 + (r/1.5\sigma_{ij})^{12}]$ is a smooth cutoff function. This potential has an attractive well that is narrower than the 6-12 form used in ref. 22, and a small potential barrier that can be considered to arise from desolvation effects upon protein binding (33). Eq. **1** produces complexes that are locally more structured (i.e., less floppy) than the original model of ref. 22.

Long-range electrostatics is an important contributor in both specific and nonspecific protein–protein interactions (1–5, 10–16). Here we use a simple Debye–Hückel-type potential, $u_{ij}^{el}(r) = q_i q_j \exp(-r/\xi)/Dr$, where q_i is the charge of residue i in Gaussian units, $\xi = 10$ Å is the Debye screening length at near-physiological salt concentrations of ~100 mM, and $D = 80$ is the dielectric constant of the water solvent. Residue charges correspond to pH 7 such that $q_i = +e$ for Lys and Arg, $-e$ for Asp and Glu, and $+0.5e$ for His, where e is the elementary charge.

The short-range interaction strengths ε_{ij} are adapted from the knowledge-based statistical contact potentials of Miyazawa and Jernigan (23). We set $\varepsilon_{ij} = \lambda(e_{ij} - e_0)$ where e_{ij} are the contact potentials. λ scales the strength of the short-range interactions compared with the physical electrostatic interactions, and e_0 balances the preference of residue–residue interactions relative to residue–solvent interactions. These two parameters are fitted against the second virial coefficient of lysozyme and the binding affinity of the ubiquitin-CUE complex (22), resulting in $\lambda = 0.416$ and $e_0 = 1.0\ k_BT$. The interaction radii, σ_{ij}, are given by $\sigma_{ij} = 1.15(\sigma_i + \sigma_j)/2$ where σ_i is the van der Waals diameter of a residue i, which is calculated from the van der Waals volume by assuming a spherical shape for the residue (22).

Simulation Method. Protein coordinates are taken from the Protein Data Bank [ID codes 1VRC (24), 1J6T (25), and 3EZA (26) for IIA^{Man}-HPr, IIA^{Mtl}-HPr, and EIN-HPr, respectively]. The protein backbone is unchanged upon complexation (24–26, 34). To obtain equilibrium properties, we perform REMC simulations on protein complexes in cubic boxes with lengths from 200 to 1,500 Å with periodic boundary conditions. Each protein is allowed to translate and rotate. To enhance equilibrium sampling at room temperature, 300 K, 20 replicas are used in the simulations at temperatures ranging from 240 to 420 K. Acceptance rates for replica exchanges varied between 40% and 70%. After equilibration for ~10^7 MC steps, ~10^8 MC steps are performed for data acquisition, saving ~10^4 configurations for analysis.

DRMS Calculation. For structure comparisons, we use the DRMS metric, $\mathrm{DRMS} = N^{-1} \Sigma_{(i,j)} |d_{ij}^{sim} - d_{ij}^{exp}|$, where N is the number of distinct residue pairs (i, j), and d_{ij}^{sim} and d_{ij}^{exp} are the distance matrices from the simulated and experimental structures, respectively.

Clustering and Structure Optimization. Simulated structures are first energy-minimized and then clustered. PRE rates for each cluster are calculated and reweighted to fit the experimental PRE rates. For this purpose, the PRE *Q*-factor (27), defined as $Q = [\Sigma_i\{\Gamma_2^{obs}(i) - \Gamma_2^{sim}(i)\}^2/\Sigma_i\Gamma_2^{obs}(i)^2]^{1/2}$, is minimized by simulated annealing, where $\Gamma_2^{obs}(i)$ and $\Gamma_2^{sim}(i)$ are the experimental and simulated PRE rates of residue i. Note that we fit the absolute PRE signal, without rescaling, but with the possibility of a small shift in PRE = 0. We also include in the fitting procedure the PRE signal derived from the experimental structure of the specific complex.

To determine the population and structures of transient encounter complexes, we minimize the PRE *Q*-factor by varying the population of transient encounter complexes, p_{minor}, and the number of clusters, N_c. The optimization procedure is as follows: (*i*) We first choose p_{minor} between 0% and 20%. (*ii*) For a given p_{minor}, we determine the cluster among the total number of clusters, N_{tot}, that gives the minimum PRE *Q*-factor, yielding

$N_c = 1$. (*iii*) Next, we minimize the PRE *Q*-factor by adding another cluster from the remaining $N_{tot} - 1$ clusters in the fitting procedure. Note that the relative weights among $N_c = 2$ clusters are optimized with the constraint of a constant combined population, p_{minor}. (*iv*) We repeat the procedure until the PRE *Q*-factor reaches a plateau as a function of the number of clusters, N_c. This procedure generates a unique set of clusters, which, however, may not necessarily be the global optimum for a given number of clusters. Note that for large numbers of clusters, the solution becomes increasingly degenerate, thus yielding multiple sets of clusters with nearly identical PRE profiles.

ACKNOWLEDGMENTS. This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

1. Schreiber G, Fersht AR (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* 3:427–431.
2. Vijayakumar M, *et al.* (1998) Electrostatic enhancement of diffusion-controlled protein-protein association: Comparison of theory and experiment on barnase and barstar. *J Mol Biol* 278:1015–1024.
3. Selzer T, Albeck S, Schreiber G (2000) Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* 7:537–541.
4. Tang C, Iwahara J, Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* 444:383–386.
5. Suh JY, Tang C, Clore GM (2007) Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement. *J Am Chem Soc* 129:12954–12955.
6. Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449:1078–1082.
7. Volkov AN, Worrall JAR, Holtzmann E, Ubbink M (2006) Solution structure and dynamics of the complex between cytochrome *c* and cytochrome *c* peroxidase determined by paramagnetic NMR. *Proc Natl Acad Sci USA* 103:18945–18950.
8. Hulsker R, Baranova MV, Bullerjahn GS, Ubbink M (2008) Dynamics in the transient complex of plastocyanin-cytochrome *f* from *Prochlorothrix hollandica*. *J Am Chem Soc* 130:1985–1991.
9. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–1025.
10. Northrup SH, Boles JO, Reynolds JCL (1988) Brownian dynamics of cytochrome *c* and cytochrome *c* peroxidase association. *Science* 241:67–70.
11. Sheinerman FB, Norel R, Honig B (2000) Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 10:153–159.
12. Miyashita O, Onuchic JN, Okamura MY (2004) Transition state and encounter complex for fast association of cytochrome c_2 with bacterial reaction center. *Proc Natl Acad Sci USA* 101:16174–16179.
13. Bui JM, McCammon JA (2006) Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci USA* 103:15451–15456.
14. Spaar A, Dammer C, Gabdoulline RR, Wade RC, Helms V (2006) Diffusional encounter of barnase and barstar. *Biophys J* 90:1913–1924.
15. Alsallaq R, Zhou HX (2007) Energy landscape and transition state of protein-protein association. *Biophys J* 92:1486–1502.
16. Alsallaq R, Zhou HX (2008) Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins* 71:320–335.
17. Zhou HX, Szabo A (2004) Enhancement of association rates by nonspecific binding to DNA and cell membranes. *Phys Rev Lett* 93:178101.
18. Prag G, *et al.* (2007) The Vps27/Hse1 complex is a GAT domain-based scaffold for ubiquitin-dependent sorting. *Develop Cell* 12:973–986.
19. Iwahara J, Clore GM (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 440:1227–1230.
20. Iwahara J, Zweckstetter M, Clore GM (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc Natl Acad Sci USA* 103:15062–15067.
21. Tang C, Ghirlando R, Clore GM (2008) Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J Am Chem Soc* 130:4048–4056.
22. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. *J Mol Biol* 375:1416–1433.
23. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term for simulation and threading. *J Mol Biol* 256:623–644.
24. Williams DC, Cai M, Suh YJ, Peterkofsky A, Clore GM (2005) Solution NMR structure of the 48-kDa $IIA^{Mannose}$-HPr complex of the *Escherichia coli* mannose phosphotransferase system. *J Biol Chem* 280:20775–20784.
25. Cornilescu G, *et al.* (2002) Solution structure of the phosphoryl transfer complex between the cytoplasmic A domain of the mannitol transporter $II^{Mannitol}$ and HPr of the *Escherichia coli* phosphotransferase system. *J Biol Chem* 277:42289–42298.
26. Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM (1999) Solution structure of the 40,000 M_r phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat Struct Biol* 6:166–173.
27. Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against ^{1}H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126:5879–5896.
28. Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* 326:1239–1259.
29. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 101:511–516.
30. Kang SA, Crane BR (2005) Effects of interface mutations on association modes and electron-transfer rates between proteins. *Proc Natl Acad Sci USA* 102:15465–15470.
31. Keskin O, Nussinov R (2007) Similar binding sites and different partners: Implications to shared proteins in cellular pathways. *Structure* 15:341–354.
32. Hu J, *et al.* (2008) Solution NMR structures of productive and non-productive complexes between the A and B domains of the cytoplasmic subunit of the mannose transport of the *Escherichia coli* phosphotransferase system. *J Biol Chem* 283:11024–11037.
33. Karanicolas J, Brooks CL (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Prot Sci* 11:2351–2361.
34. Garrett DS, *et al.* (1997) Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry* 36:2517–2530.