Identifying hydrophobic protein patches to inform protein interaction interfaces

Nicholas B. Rego[a], Erte Xi[b], and Amish J. Patel[a,b,1]

[a]Biochemistry and Molecular Biophysics Graduate Group, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104

Edited by Ronald M. Levy, Temple University, Philadelphia, PA, and accepted by Editorial Board Member Peter J. Rossky December 13, 2020 (received for review August 30, 2020)

Interactions between proteins lie at the heart of numerous biological processes and are essential for the proper functioning of the cell. Although the importance of hydrophobic residues in driving protein interactions is universally accepted, a characterization of protein hydrophobicity, which informs its interactions, has remained elusive. The challenge lies in capturing the collective response of the protein hydration waters to the nanoscale chemical and topographical protein patterns, which determine protein hydrophobicity. To address this challenge, here, we employ specialized molecular simulations wherein water molecules are systematically displaced from the protein hydration shell; by identifying protein regions that relinquish their waters more readily than others, we are then able to uncover the most hydrophobic protein patches. Surprisingly, such patches contain a large fraction of polar/charged atoms and have chemical compositions that are similar to the more hydrophilic protein patches. Importantly, we also find a striking correspondence between the most hydrophobic protein patches and regions that mediate protein interactions. Our work thus establishes a computational framework for characterizing the emergent hydrophobicity of amphiphilic solutes, such as proteins, which display nanoscale heterogeneity, and for uncovering their interaction interfaces.

hydrophobicity | proteins | PPI | hydration | dewetting

Protein–protein interactions play a crucial role in numerous biological processes, ranging from signal transduction and immune response to protein aggregation and phase behavior (1–3). Consequently, being able to understand, predict, and modulate protein interactions has important implications for understanding cellular processes and mitigating the progression of disease (4, 5). A necessary first step toward this ambitious goal is uncovering the interfaces through which proteins interact (6–8). In principle, identifying hydrophobic protein regions, which interact weakly with water, should be a promising strategy for uncovering protein interaction interfaces (9, 10). Indeed, the release of weakly interacting hydration waters from hydrophobic regions can drive protein interactions, as well as other aqueous assemblies (11–13). However, even when the structure of a protein is available at atomistic resolution, it is challenging to identify its hydrophobic patches because they are not uniformly nonpolar, but display variations in polarity and charge at the nanoscale. Moreover, the emergent hydrophobicity of a protein patch stems from the collective response of protein hydration waters to the nanoscale chemical and topographical patterns displayed by the patch (14–20) and cannot be captured by simply counting the number of nonpolar groups in the patch, or even through more involved additive approaches, such as hydropathy scales or surface-area models (21–28).

To address this challenge, we build upon seminal theoretical advances and molecular simulation studies, which have shown that near a hydrophobic surface, it is easier to disrupt surface–water interactions and form interfacial cavities (29–34). To uncover protein regions that have the weakest interactions with water, here, we employ specialized molecular simulations, wherein protein–water interactions are disrupted by systematically displacing water molecules from the protein hydration shell (35–37). By identifying the protein patches that nucleate cavities most readily in our simulations, we are then able to uncover the most hydrophobic protein regions. Interestingly, we find that both hydrophobic and hydrophilic protein patches are highly heterogeneous and contain comparable numbers of nonpolar and polar atoms. Our results thus highlight the nontrivial relationship between the chemical composition of protein patches and their emergent hydrophobicity (24–26), and further emphasize the importance of accounting for the collective solvent response in characterizing protein hydrophobicity (16). We then interrogate whether the most hydrophobic protein patches, which nucleate cavities readily, are also likely to mediate protein interactions. Across five proteins that participate in either homodimer or heterodimer formation, we find that roughly 60 to 70% of interfacial contacts and only about 10 to 20% of noncontacts nucleate cavities. Our work thus provides a versatile computational framework for characterizing hydrophobicity and uncovering interaction interfaces of not just proteins, but also of other complex amphiphilic solutes, such as cavitands, dendrimers, and patchy nanoparticles (38–41).

Systematically Disrupting Protein–Water Interactions

We illustrate our approach using the protein thymidylate synthase (TS), which forms a homodimer and is involved in DNA

Significance

Hydrophobicity drives diverse aqueous assemblies, including the interactions between proteins. Conventional wisdom stipulates that the hydrophobicity of a surface must be inversely related to its polarity. By using specialized molecular simulations to characterize protein hydrophobicity, here, we challenge this notion and demonstrate that polar/charged atoms comprise a substantial fraction of hydrophobic protein patches. Moreover, hydrophobic and hydrophilic protein patches have surprisingly similar chemical compositions. We further find that the most hydrophobic protein patches also tend to mediate protein interactions. Our work thus establishes a computational framework for characterizing the hydrophobicity of complex solutes, such as proteins, and for uncovering their interaction interfaces.

Author contributions: N.B.R., E.X., and A.J.P. designed research; N.B.R. performed research; N.B.R. analyzed data; and N.B.R. and A.J.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. R.M.L. is a Guest Editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: amish.patel@seas.upenn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2018234118/-/DCSupplemental.

Published February 1, 2021.

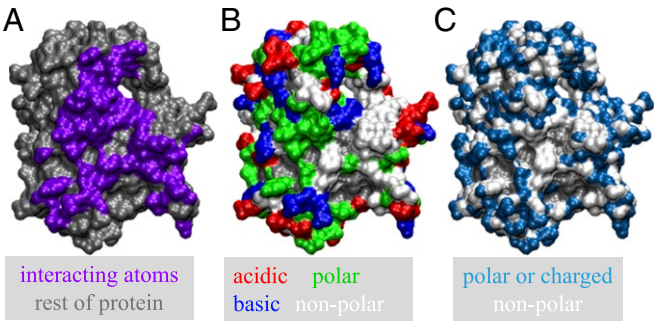

Fig. 1. Identifying protein patches that mediate its interactions. (*A*) The atoms of TS that participate in the formation of an obligate homodimer are projected onto a TS monomer. Although the knowledge of such interaction interfaces is desirable, it is unavailable for most proteins. (*B* and *C*) Protein structures provide access to the nanoscale chemical and topographical patterns displayed by the protein; however, using this information to predict the protein patches that mediate its interactions is nontrivial. Such patterns are shown here for the TS protein, wherein each residue is colored according to its overall chemistry (*B*) and each atom is colored using the Kapcha–Rossky classification (42) (*C*).

synthesis (43); protein atoms that participate in dimer formation are shown in Fig. 1*A*. Visualizations of the protein, colored according to residue (Fig. 1*B*) or atomic (Fig. 1*C*) chemistry, highlight the nanoscopic chemical and topographical patterns displayed by the TS protein. Comparing these visualizations to the protein-interaction interface (Fig. 1*A*) also illustrates the lack of any obvious distinguishing characteristics, which can be used to identify the interaction interface.

To interrogate whether the most hydrophobic regions of the TS protein are also likely to mediate its interactions, we must first characterize how (un)favorably different protein regions interact with water. To this end, we perform all-atom, explicit-solvent, molecular dynamics simulations, wherein an unfavorable biasing potential, ϕN_v, is applied to systematically disrupt protein–water interactions (44, 45); ϕ represents the potential strength, and N_v is the number of (coarse-grained) waters in the protein hydration shell, v. To ensure that v conforms to the rugged protein surface, we peg spherical subvolumes to every heavy atom on the protein surface and define v to be the union of all such subvolumes (Fig. 2*A*); by choosing the subvolume radius, R_v, to be 0.6 nm, we include only the first hydration shell waters in v. We refer to such biased simulations as "ϕ-ensemble simulations" and describe them in detail in *SI Appendix*.

As the strength of the potential, ϕ, is increased, waters are displaced from v, and the average number of waters, $\langle N_v \rangle_\phi$, in the protein hydration shell decreases. The response of the TS hydration waters to the potential is shown in Fig. 2*B*; as ϕ is increased, $\langle N_v \rangle_\phi$ decreases sigmoidally. Correspondingly, the susceptibility, $\chi_v \equiv -\partial \langle N_v \rangle_\phi / \partial(\beta\phi)$, displays a marked peak (Fig. 2*C*), signifying that the TS protein hydration shell undergoes a collective dewetting transition in response to the unfavorable potential. By studying proteins with a variety of sizes, shapes, chemistries, and biological functions, we recently showed that collective dewetting is a generic feature of protein hydration shells (46). For the TS protein, the peak in susceptibility occurs at $\beta\phi^* = 2.16$, where $\beta^{-1} \equiv k_B T$, k_B is the Boltzmann constant, and T is temperature. The potential strengths for which χ_v is half its maximum value, $\beta\phi^{(-)} = 1.6$ and $\beta\phi^{(+)} = 3.4$, are also shown in Fig. 2*C* and correspond roughly to the onset and the end of dewetting, respectively. The disruption of protein–water interactions by the unfavorable potential, and the resultant dewetting, lead to the formation of cavities in the protein hydration shell; dewetting is not uniform, but instead manifests as wet and dry patches on the protein surface. Such patches, observed in simulations with $\beta\phi = 1.8,\ 2.4$, and 3, are shown in Fig. 2*D*.

Identifying Hydrophobic Protein Patches

The cavities that form in response to an unfavorable potential (Fig. 2*D*) undergo substantial fluctuations in their size, shape, and location over the course of a ϕ-ensemble simulation. To identify protein regions that nucleate cavities, we thus estimate the average water density in different parts of the protein hydration shell. In particular, we estimate the average number of waters, $\langle n_i \rangle_\phi$, in the spherical subvolume centered on every protein surface heavy atom i and normalize it by the corresponding unbiased average, $\langle n_i \rangle_0$. For the TS protein, a map of the normalized local water density $\langle \rho_i \rangle_\phi \equiv \langle n_i \rangle_\phi / \langle n_i \rangle_0$ is shown in Fig. 3*A* for $\beta\phi = 2.4$. Protein atoms whose $\langle \rho_i \rangle_\phi$ values fall below a certain threshold, $s = 0.5$—i.e., atoms that lose at least half of their hydration waters—are then classified as being dewetted; *SI Appendix*, Fig. S7. In Fig. 3*B*, protein atoms that are dewetted at $\beta\phi = 2.4$ are shown in orange, whereas those that remain wet are shown in gray. Protein regions that dewet do so because, overall, they have weaker interactions with water than the regions

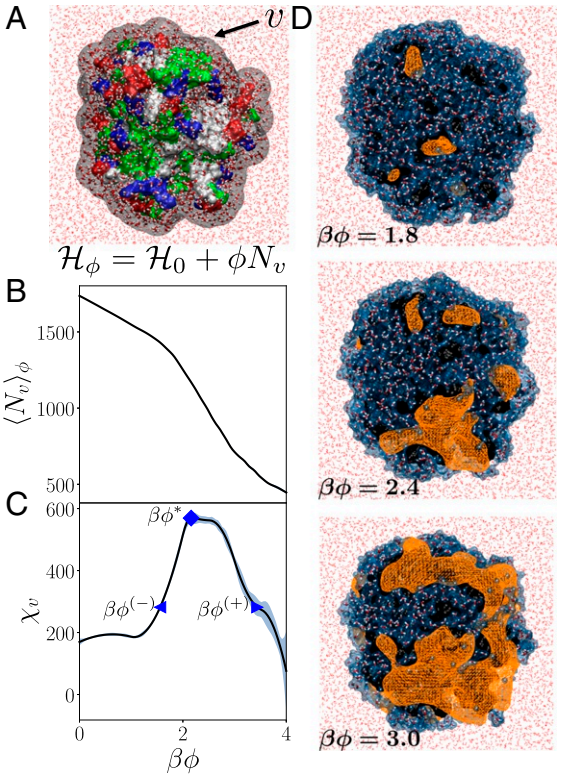

Fig. 2. Disrupting protein–water interactions using ϕ-ensemble simulations. (*A*) The hydration shell, *v*, of the TS protein is shown (transparent gray). The protein surface is colored by residue (as in Fig. 1*B*), the waters in *v* are shown in licorice, and the rest are shown as lines. In ϕ-ensemble simulations, an unfavorable biasing potential, ϕN_v, is applied to the N_v waters in *v*. (*B*) As the strength of the potential ϕ is increased, the average number of waters, $\langle N_v \rangle_\phi$, decreases in a sigmoidal manner. (*C*) The corresponding susceptibility, $\chi_v = -\partial \langle N_v \rangle_\phi / \partial(\beta\phi)$, displays a peak at $\beta\phi^* = 2.16$ (diamond), highlighting that dewetting of the protein hydration shell is collective. The potential strengths that mark the onset $\phi^{(-)}$ (left triangle) and the end $\phi^{(+)}$ (right triangle) of the peak in χ_v are also shown. (*D*) Simulation snapshots are shown for ϕ ensembles corresponding to $\beta\phi = 1.8$ (*Top*), $\beta\phi = 2.4$ (*Middle*), and $\beta\phi = 3.0$ (*Bottom*). Protein atoms are shown in surface representation (black), hydration waters as licorice, and the rest as lines. The waters in *v* are surrounded by a blue mesh, whereas cavities are shown using an orange mesh.

that remain wet; we thus classify such regions as being more hydrophobic, regardless of their chemical identities (e.g., their polarity or charge).

To what extent is the hydrophobicity of a protein patch determined by the chemistry of its constituent atoms? In other words, is a hydrophobic protein patch, which dewets in response to an unfavorable potential, composed primarily of nonpolar atoms? Are the (hydrophilic) protein atoms that remain wet primarily polar or charged? To answer these questions, we follow the Kapcha–Rossky prescription and classify protein atoms as nonpolar or polar/charged based on their partial charges, $\{q_i\}$, in the optimized potentials for liquid simulations force field; a protein atom i is deemed nonpolar if $|q_i| < 0.25$, and polar/charged otherwise (42). As shown in Fig. 3*C*, roughly one-fifth of the protein surface is dewetted (orange) at $\beta\phi = 2.4$, and the rest remains wet (gray). Interestingly, the hydrophobic protein patch, which dewets at $\beta\phi = 2.4$, is not predominantly nonpolar; instead, nearly 40% of its atoms are either polar or charged. Moreover, the rest of the protein, which remains wet, is not predominantly polar/charged either; nearly 40% of its atoms are nonpolar! Although the fraction of dewetted protein atoms increases with ϕ, the relatively hydrophobic (dewetted) and hydrophilic (wet) parts of the protein remain highly heterogeneous with chemical compositions that are remarkably similar; Fig. 3*D* and Movie S1. These findings hold true for all of the proteins we study here (Movies S1–S5) and stem from the fact that when certain nonpolar protein regions dewet, neighboring protein regions, which are often polar, become much more susceptible to dewetting. Collectively, these results highlight that, in contrast with homogeneous surfaces, the hydrophobicity of a heterogeneous protein patch is only poorly correlated with the polarity of its constituent atoms; instead, predicting protein hydrophobicity requires an accurate accounting of the collective solvent response to the chemical and topographical protein context.

Comparing Dewetted Atoms and Protein Contacts

Can such a holistic characterization of the hydrophobicity of a protein be used to inform its interaction interfaces? To answer this question, we compare hydrophobic regions of the TS protein, which dewet in ϕ-ensemble simulations, against the TS interaction interface. The 210 interaction-interface atoms (or "contacts"), shown in Fig. 4*A* (purple), constitute 15.7% of the protein surface and were determined from the crystal structure of the TS homodimer, as described in *SI Appendix*, Fig. S2. Similarly, the 249 protein atoms which dewet at $\beta\phi = 2.4$ are shown in Fig. 4*B* (orange). From these snapshots, it is clear that there is substantial overlap between the two sets of atoms. To further quantify this correspondence, we classify protein surface heavy atoms into four categories, which are illustrated schematically in Fig. 4*C*: protein contacts that dewet are true positives (TPs; pink), whereas contacts that remain wet are false negatives (FNs; dark purple); similarly, noncontacts that dewet are false positives (FPs; dark orange), whereas noncontacts that remain wet are true negatives (TNs; gray). The TS protein atoms, classified accordingly, are shown in Fig. 4*C* for three different ϕ ensembles. Snapshots for other ϕ ensembles and those showing the rest of the protein are included in Movie S1. For ϕ near $\phi^{(-)}$, very few atoms are dewetted, and a substantial fraction of the interaction interface remains wet; thus, numerous FNs are seen at $\beta\phi = 1.8$. Conversely, for ϕ near $\phi^{(+)}$, much of the protein surface is dewetted; consequently, many of the dewetted atoms do not belong to the interaction interface, and several FPs are seen at $\beta\phi = 3$. Intermediate values of ϕ in the vicinity of ϕ^* provide the optimal balance, dewetting most of the interaction-interface atoms (TPs), while minimizing the FNs and FPs.

To assess the performance of the different ϕ ensembles more quantitatively, we plot the fraction of protein contacts that are dewetted (TP rate; TPR), as well as the fraction of noncontacts that are dewetted (FP rate; FPR) as functions of ϕ (Fig. 4*D*). As ϕ is increased from roughly $\beta\phi^{(-)} = 1.6$ to $\beta\phi^* \approx 2.2$, TPR increases sharply; as more protein atoms dewet, an increasingly larger fraction of the interaction interface is identified. Importantly, FPR remains close to zero throughout this ϕ range, suggesting that the majority of the protein atoms which dewet in this range indeed belong to the interaction interface (Movie S1). For example, 87% of the protein atoms that dewet at $\beta\phi = 1.96$ are contacts. As ϕ is increased further, TPR continues to increase, but does so more gradually. Moreover, FPR increases rapidly from roughly $\beta\phi^* = 2.2$ to $\beta\phi^{(+)} \approx 3.4$. Thus, an increasing fraction of protein atoms that dewet at these higher ϕ values are noncontacts. To better visualize the trade-off between TPR and FPR with increasing ϕ, we plot them against one another in a receiver operating characteristic (ROC) curve (Fig. 4*E*). For $\phi^{(-)} < \phi < \phi^*$, TPR increases, while FPR remains small; the ROC curve thus increases sharply at first. As ϕ is increased further, FPR also starts to increase, so the ROC curve climbs less sharply.

The perfect correspondence, wherein all of the contacts (and only the contacts) are dewetted, corresponds to TPR = 1 and

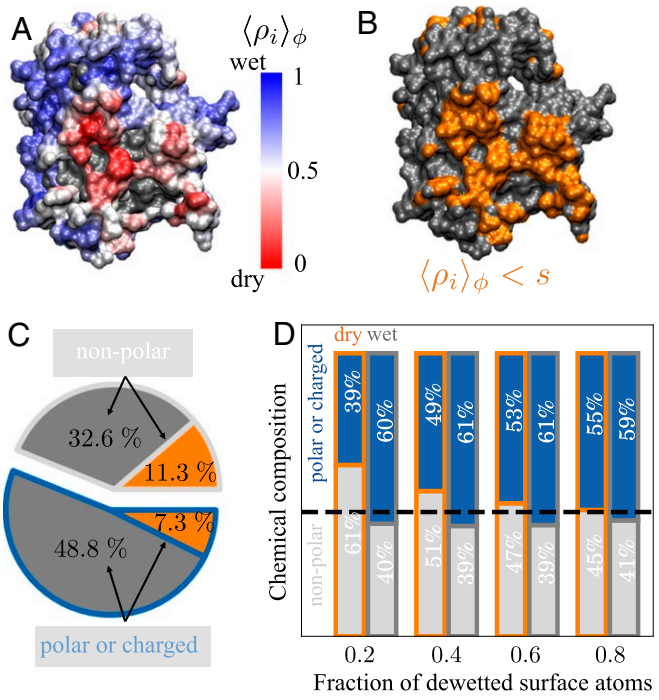

Fig. 3. Uncovering hydrophobic protein patches by identifying regions that dewet in ϕ-ensemble simulations. (*A*) Snapshot of the TS protein is shown with every protein heavy atom, *i*, colored according to the average water density, $\langle\rho_i\rangle_\phi$, in its hydration shell at $\beta\phi = 2.4$; dewetted atoms are colored red, whereas atoms that remain hydrated are colored blue. (*B*) Protein atoms for which $\langle\rho_i\rangle_\phi$ falls below a threshold, $s = 0.5$, are considered to be dewetted and are shown in orange; the rest are shown in gray. (*C*) Protein atoms are categorized according to whether they dewet (orange fill) or not (gray fill) at $\beta\phi = 2.4$, as well as whether they are nonpolar (white outline) or polar/charged (blue outline) according to the Kapcha–Rossky classification (42). Interestingly, only 60% of the dewetted protein atoms (orange fill) are nonpolar (white outline), whereas the remaining 40% of the atoms are polar/charged (blue outline). (*D*) As ϕ is increased and a larger fraction of the protein surface dewets, the hydrophobic (dewetted) and hydrophilic (wet) protein regions remain heterogeneous and have chemical compositions that are remarkably similar; the dashed line represents the overall composition of the protein surface.

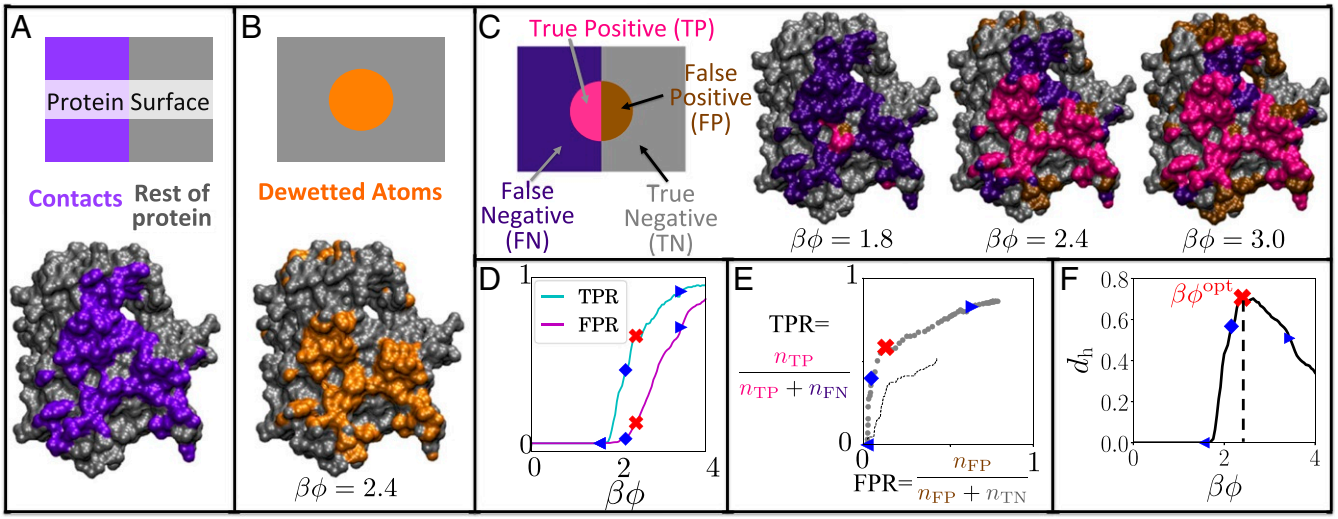

Fig. 4. Protein hydrophobicity informs protein-interaction interfaces. (*A*) Schematic (*Upper*) and TS protein structure (*Lower*) highlighting protein atoms (purple) which participate in the formation of the TS homodimer and those that don't (gray). (*B*) Schematic (*Upper*) and TS protein structure (*Lower*) showing protein atoms that are dewetted at $\beta\phi = 2.4$ (orange) and those that remain wet (gray). (*C*) By comparing dewetted protein atoms against those belonging to the interaction interface (contacts), we identify protein contacts that dewet (TP; pink) and those that remain wet (FN; dark purple), as well as noncontacts that dewet (FP; dark orange) and ones that stay wet (TN; gray). Schematic (*Left*) and TS protein structures (*Right*) illustrate such a comparison for $\beta\phi = 1.8$, 2.4, and 3. Very few atoms are dewetted at $\beta\phi = 1.8$; consequently, much of the interaction interface remains wet (dark purple). In contrast, much of the protein surface is dewetted at $\beta\phi = 3$, including several noncontacts (dark orange). The right balance between FNs and FPs is achieved at $\beta\phi = 2.4$, where most of the contacts are dewetted, and most of the noncontacts remain wet. (*D*) Both the fraction of contacts that dewet (TPR) and the fraction of noncontacts that dewet (FPR) display a sigmoidal increase with increasing ϕ. (*E*) The ROC curve illustrates the variation of TPR with FPR; symbols correspond to ϕ-ensemble simulations, whereas the dashed line corresponds to nonpolar clusters (see *Comparing Dewetted Atoms and Protein Contacts*). (*F*) The harmonic average, d_h, of TPR and $1 - \mathrm{FPR}$, shown as a function of ϕ, displays a peak at $\beta\phi^{opt} = 2.4$ with a peak value, $d_h^{opt} = 0.71$. In *D*–*F*, the left triangle, diamond, cross, and right triangle symbols correspond to ϕ values of $\phi^{(-)}$, ϕ^*, ϕ^{opt}, and $\phi^{(+)}$, respectively.

$\mathrm{FPR} = 0$. To quantify proximity to this ideal, we estimate the harmonic average, d_h, of TPR and $1 - \mathrm{FPR}$ for every ϕ ensemble. This choice ensures that only when both $\mathrm{TPR} \to 1$ and $\mathrm{FPR} \to 0$ does $d_h \to 1$; in contrast, when either $\mathrm{TPR} \to 0$ or $\mathrm{FPR} \to 1$, $d_h \to 0$. In Fig. 4*F*, we plot d_h as a function of ϕ for the TS protein. For $\phi < \phi^{(-)}$, $d_h = 0$ because $\mathrm{TPR} = 0$, i.e., unbiased simulations or those with low ϕ values do not nucleate cavities, and are thus ill-suited to informing the interaction interface. Conversely, for $\phi > \phi^{(+)}$, much of the protein is dewetted, and d_h decreases sharply as FPR approaches one. Intermediate ϕ values provide the optimal trade-off with a potential strength of $\beta\phi^{opt} = 2.4$, resulting in the maximum d_h value of $d_h^{opt} = 0.71$. In the judiciously chosen ϕ^{opt} ensemble, nearly 60% of the protein contacts and only 10% of the noncontacts dewet, and roughly 50% of the dewetted atoms belong to the interaction interface. Collectively, these numbers and similar such metrics included in *SI Appendix*, Tables S1–S3 quantify the extent to which the hydrophobicity of a protein informs its interaction interface.

To quantify how simpler approaches, which use protein-surface chemistry alone, might fare in informing interaction interfaces (relative to ϕ-ensemble simulations), we construct a family of increasingly larger nonpolar clusters, as detailed in *SI Appendix*, and compare them against the interaction-interface atoms. The corresponding ROC curves, shown in Fig. 4*E* for the TS protein (dashed line) and in *SI Appendix*, Figs. S9*F*–S13*F* for other proteins studied here, highlight that while ϕ-ensemble simulations consistently inform interaction interfaces, the correlation between the nonpolar clusters and protein contacts is relatively poor and quite variable (47).

Homodimer- and Heterodimer-Forming Proteins

Can protein hydrophobicity, characterized by using ϕ-ensemble simulations, be used to inform the interaction interfaces of other proteins? To address this question, we investigate four additional proteins, two of which form homodimers and two that form heterodimers. The homodimer-forming proteins are 1) the lectin domain of rat mannose-binding protein (MBP)—a protein that contributes to the innate immune response by recognizing sugars on bacterial cell walls (48)—and 2) a dimer of the bee-venom protein melittin (MLT), which can interact with another dimer to form a tetramer (49). Compared with proteins that participate in obligate homodimeric interactions, the interactions between proteins that form heterodimers tend be weaker, featuring interfaces that are smaller and more amphiphilic (50–52). As a result, identifying heterodimeric interfaces tends to be more challenging. Nevertheless, being able to identify such interaction interfaces is important due to their prevalence in cell signaling. We thus study the heterodimer-forming proteins: 1) the mouse double minute 2 homolog (MDM2)—an inhibitor of the tumor-suppressor protein, p53, whose mutations have been implicated in a number of cancers (53)—and 2) ubiquitin (UBQ)—a highly conserved protein known to form heterodimeric interactions with a variety of binding partners (54). To determine its interaction interface, we thus use structures of ubiquitin in complex with four of its binding partners (*SI Appendix*, *SI Text* and Fig. S3).

We performed and analyzed ϕ-ensemble simulations for MBP, MLT, MDM2, and UBQ and compared the protein atoms that dewet against the experimentally determined interaction-interface atoms. The results are qualitatively similar to those shown in Figs. 2–4 for the TS protein and are summarized in Fig. 5 with the underlying analysis included in *SI Appendix*, Figs. S9–S13. The correspondence between dewetted protein atoms and known contacts, as quantified by the optimal d_h scores, is shown in Fig. 5*A*, and for all of the proteins studied here, $d_h^{opt} \approx 0.71 \pm 0.07$ (*SI Appendix*, Table S1). The optimal ϕ ensembles are also shown in Fig. 5*A* and highlight that for every

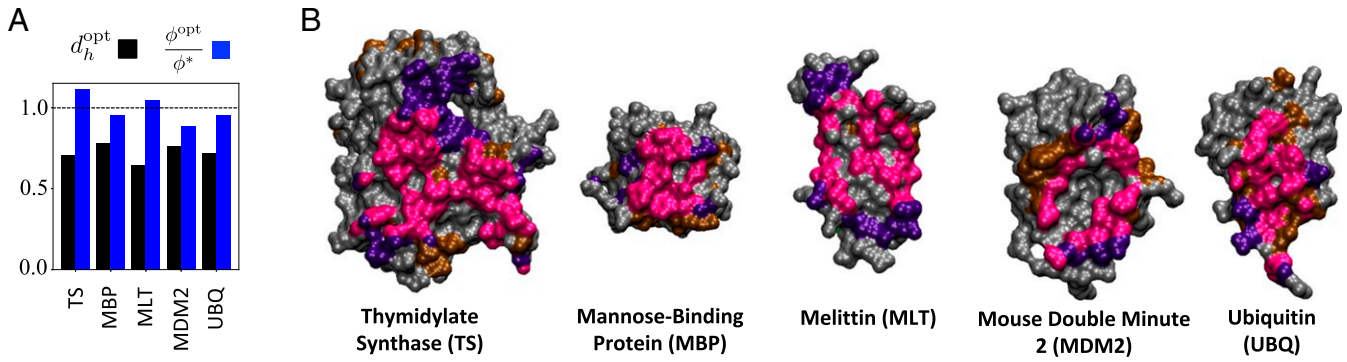

Fig. 5. Proteins that participate in the formation of homodimers and heterodimers. (*A*) The correspondence between hydrophobic protein patches that dewet in the ϕ^{opt} ensemble and protein-interaction interfaces is quantified by using d_h^{opt}; for the five proteins studied here, we find $d_h^{\mathrm{opt}} \approx 0.7$. In each case, we further find that $\phi^{\mathrm{opt}} \approx \phi^*$; the optimal potential strength for uncovering the interaction interface is similar to the potential strength to which the protein hydration waters are most susceptible. (*B*) Protein atoms that dewet in the ϕ^{opt} ensemble are compared against protein contacts for the self-interacting MBP and MLT dimer, as well as for the MDM2 protein and UBQ, which interact with other proteins to form heterodimers. The color scheme used here is the same as in Fig. 4*C*. In each case, the center of the interaction interface tends to dewet (TPs; pink), whereas the periphery of the interface features FNs and FPs.

protein studied here, $\phi^{\mathrm{opt}} \approx \phi^*$, suggesting that ϕ^* provides an excellent first guess for ϕ^{opt}.

In Fig. 5*B*, protein atoms dewetted in the ϕ^{opt} ensemble are compared against the interaction-interface atoms for TS, MBP, MLT, MDM2, and UBQ. These visualizations highlight that the center of the interface is dewetted in all cases and that contacts which do not dewet (FNs) tend to be concentrated along the interface edges. The presence of FNs along the interface periphery is consistent with the notion that peripheral regions, which are partially exposed to water in the bound state, ought to interact more favorably with water than the center of the interface (50), and should thus be less likely to dewet. Snapshots of protein regions that are not visible in Fig. 5*B* are included in *SI Appendix*, Figs. S9*K*–S13*K* and Movies S1–S5. These snapshots highlight that for the five proteins studied here, noncontacts that dewet (FPs) tend to either extend outward from the interaction interface (i.e., the interface uses only part of the hydrophobic patch), or they are small, isolated clusters. We note that, although a single, contiguous dewetted patch is observed for the relatively simple proteins studied here, for more complex proteins, which employ disparate interaction sites (1), we expect that multiple different protein regions would dewet. By uncovering the hydrophobic patches on such proteins, as well as the order in which they dewet, our approach could inform the formation of multimeric protein assemblies (55, 56).

Conclusions and Outlook

Using specialized molecular simulations, wherein an unfavorable potential is employed to systematically displace protein hydration waters and nucleate cavities, here, we identify the most hydrophobic protein regions, which have the weakest interactions with water. Interestingly, we find that polar/charged atoms comprise a substantial fraction of hydrophobic protein patches. Moreover, such hydrophobic patches and the more hydrophilic protein patches have surprisingly similar chemical compositions. These findings suggest that, unlike homogeneous surfaces, the hydrophobicity of a heterogeneous protein patch cannot be inferred from the polarity/charge of its constituent atoms alone. They further highlight that an appropriate accounting of the collective response of the protein hydration waters to the nanoscale chemical and topographical patterns on the protein surface is necessary for an accurate characterization of protein hydrophobicity (22–28).

Importantly, we further find that the most hydrophobic protein patches, which dewet readily in our specialized simulations, also correspond closely with the protein regions that mediate its interactions. In particular, for the five proteins studied here, the most hydrophobic protein patch, which dewets in response to a judiciously chosen unfavorable potential, contains 60 to 75% of the protein contacts and only 10 to 25% of noncontacts. Such a correspondence between hydrophobic patches and interaction interfaces is even more remarkable than might appear at first glance. In particular, when one protein binds another, protein–water interactions are replaced with direct interactions between proteins (56). Thus, the fact that patches with weak protein–water interactions can inform protein-interaction interfaces, without any consideration of the direct interactions between proteins, is particularly noteworthy.

Integrating our framework for characterizing protein hydrophobicity with prescriptions for capturing direct protein–protein interactions could thus yield a more complete characterization of protein interactions. We expect such a holistic approach to be particularly fruitful when proteins are driven together by strong direct interactions, such as complementary electrostatic interactions between dipolar or charged groups. For example, scoring functions that complement protein docking (57, 58) tend to capture the direct interactions between proteins quite well, but their use of additive approximations to account for protein hydrophobicity (e.g., hydropathy scales or surface-area models) introduces substantial inaccuracies and limits their overall utility (58, 59). By providing a computational framework for accurately characterizing protein hydrophobicity, we hope that our work will not only lead to more accurate scoring functions, but also result in a better understanding of the complex relationship between chemical and topographical protein patterns and their emergent hydrophobicities. Alternatively, identification of the putative interaction interfaces could be used as a starting point for explicit-solvent, enhanced sampling simulations of protein binding. Although computationally expensive, such simulations can provide an exquisite quantitative characterization of both the thermodynamics and kinetics of protein interactions (56, 60–63).

Undesirable interactions between proteins and their subsequent aggregation play an important role in neurodegenerative diseases and frustrate the storage and transportation of concentrated protein formulations (64, 65). By uncovering the most hydrophobic protein patches, our approach could facilitate the identification of aggregation-prone regions, as well as hot-spots residues, which make outsize contributions to binding affinity (66, 67) and are targets for enhancing protein solubility (68, 69).

Because our approach relies on capturing the collective solvent response to the protein surface, it can be readily generalized to investigate how the presence of cosolutes influences protein hydrophobicity and the corresponding interaction interfaces (70, 71). Similarly, our approach should also be useful in informing the interactions of nanostructured solutes other than proteins (38–41).

Data Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. A.J.P. was supported by NSF Grants CBET-1652646, CHE-1665339, and DMR-1720530; Alfred P. Sloan Research Foundation Award FG-2017-9406; and Camille & Henry Dreyfus Foundation Award TC-19-033. N.B.R. was supported by NSF Grants CBET-1652646 and DMR-1844514. We thank the reviewer who suggested the nonpolar clustering analysis.

1. A. S. Futran, A. J. Link, R. Seger, S. Y. Shvartsman, ERK as a model for systems biology of enzyme kinetics in cells. *Curr. Biol.* **23**, R972–R979 (2013).
2. D. Thirumalai, G. Reddy, J. E. Straub, Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.* **45**, 83–92 (2012).
3. S. R. Trevino, J. M. Scholtz, C. N. Pace, Measuring and increasing protein solubility. *J. Pharmacol. Sci.* **97**, 4155–4166 (2008).
4. D. E. Scott, A. R. Bayly, C. Abell, J. Skidmore, Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* **15**, 533–550 (2016).
5. I. Petta, S. Lievens, C. Libert, J. Tavernier, K. De Bosscher, Modulation of protein-protein interactions for the development of novel therapeutics. *Mol. Ther.* **24**, 707–718 (2016).
6. L. C. Xue, D. Dobbs, A. M. J. J. Bonvin, V. Honavar, Computational prediction of protein interfaces: A review of data driven methods. *FEBS Lett.* **589**, 3516–3526 (2015).
7. F. Bai, F. Morcos, R. R. Cheng, H. Jiang, J. N. Onuchic, Elucidating the druggable interface of protein-protein interactions using fragment docking and coevolutionary analysis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E8051–E8058 (2016).
8. P. Gainza *et al.*, Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
9. L. Young, R. L. Jernigan, D. G. Covell, A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.* **3**, 717–729 (1994).
10. O. Keskin, A. Gursoy, B. Ma, R. Nussinov, Principles of protein-protein interactions: What are the preferred ways for proteins to interact?*Chem. Rev.* **108**, 1225–1244 (2008).
11. K. A. Dill, T. M. Truskett, V. Vlachy, B. Hribar-Lee, Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 173–199 (2005).
12. Y. Levy, J. N. Onuchic, Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415 (2006).
13. D. Reichmann, Y. Phillip, A. Carmi, G. Schreiber, On the contribution of water-mediated interactions to protein-complex stability. *Biochemistry* **47**, 1051–1060 (2008).
14. N. Giovambattista, C. F. Lopez, P. J. Rossky, P. G. Debenedetti, Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2274–2279 (2008).
15. P. J. Rossky, Exploring nanoscale hydrophobic hydration. *Faraday Discuss* **146**, 13–18 (2010).
16. S. N. Jamadagni, R. Godawat, S. Garde, Hydrophobicity of proteins and interfaces: Insights from density fluctuations. *Annu. Rev. Chem. Biomol. Eng.* **2**, 147–171 (2011).
17. M. Heyden, D. J. Tobias, Spatial dependence of protein-water collective hydrogen-bond dynamics. *Phys. Rev. Lett.* **111**, 218101 (2013).
18. A. C. Fogarty, E. Duboué-Dijon, F. Sterpone, J. T. Hynes, D. Laage, Biomolecular hydration dynamics: A jump model perspective. *Chem. Soc. Rev.* **42**, 5672–5683 (2013).
19. D. Cui, S. Ou, S. Patel, Protein-spanning water networks and implications for prediction of protein–protein interactions mediated through hydrophobic effects. *Proteins* **82**, 3312–3326 (2014).
20. J. Monroe *et al.*, Water structure and properties at hydrophilic and hydrophobic surfaces. *Annu. Rev. Chem. Biomol. Eng.* **11**, 523–557 (2020).
21. N. T. Southall, K. A. Dill, A. D. J. Haymet, A view of the hydrophobic effect. *J. Phys. Chem. B* **106**, 521–533 (2002).
22. J. Mittal, G. Hummer, Interfacial thermodynamics of confined water near molecularly rough surfaces. *Faraday Discuss* **146**, 341–352 (2010).
23. R. C. Harris, B. M. Pettitt, Effects of geometry and chemistry on hydrophobic solvation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14681–14686 (2014).
24. H. Acharya, S. Vembanur, S. N. Jamadagni, S. Garde, Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday Discuss* **146**, 353–365 (2010).
25. J. Wang, D. Bratko, A. Luzar, Probing surface tension additivity on chemically heterogeneous surfaces by a molecular approach. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6374–6379 (2011).
26. E. Xi *et al.*, Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13345–13350 (2017).
27. C. D. Ma, C. Wang, C. Acevedo-Vélez, S. H. Gellman, N. L. Abbott, Modulation of hydrophobic interactions by proximally immobilized ions. *Nature* **517**, 347–350 (2015).
28. J. I. Monroe, M. S. Shell, Computational discovery of chemically patterned surfaces that effect unique hydration water dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8093–8098 (2018).
29. K. Lum, D. Chandler, J. D. Weeks, Hydrophobicity at small and large length scales. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
30. J. D. Weeks, Connecting local structure to interface formation: A molecular scale van der Waals theory of nonuniform liquids. *Annu. Rev. Phys. Chem.* **53**, 533–562 (2002).
31. A. Wallqvist, E. Gallicchio, R. M. Levy, A model for studying drying at hydrophobic interfaces: Structural and thermodynamic properties. *J. Phys. Chem. B* **105**, 6745–6753 (2001).
32. R. Godawat, S. N. Jamadagni, S. Garde, Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15119–15124 (2009).
33. A. J. Patel, D. Chandler, Fluctuations of water near extended hydrophobic and hydrophilic surfaces. *J. Phys. Chem. B* **114**, 1632–1637 (2010).
34. S. Vaikuntanathan, G. Rotskoff, A. Hudson, P. L. Geissler, Necessity of capillary modes in a minimal model of nanoscale hydrophobic solvation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2224–E2230 (2016).
35. A. J. Patel *et al.*, Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *J. Phys. Chem. B* **116**, 2498–2503 (2012).
36. A. J. Patel, S. Garde, Efficient method to characterize the context-dependent hydrophobicity of proteins. *J. Phys. Chem. B* **118**, 1564–1573 (2014).
37. E. Xi, R. C. Remsing, A. J. Patel, Sparse sampling of water density fluctuations in interfacial environments. *J. Chem. Theor. Comput.* **12**, 706–713 (2016).
38. I. C. Pons-Siepermann, S. C. Glotzer, Design of patchy particles using quaternary self-assembled monolayers. *ACS Nano* **6**, 3919–3924 (2012).
39. Q. Xiao *et al.*, Janus dendrimersomes coassembled from fluorinated, hydrogenated, and hybrid Janus dendrimers as models for cell fusion and fission. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7045–E7053 (2017).
40. Z. Luo *et al.*, Determination and evaluation of the nonadditivity in wetting of molecularly heterogeneous surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25516–25523 (2019).
41. J. W. Barnett *et al.*, Spontaneous drying of non-polar deep-cavity cavitand pockets in aqueous solution. *Nat. Chem.* **12**, 589–594 (2020).
42. L. H. Kapcha, P. J. Rossky, A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* **426**, 484–498 (2014).
43. M. G. Rose, M. P. Farrell, J. C. Schmitz, Thymidylate synthase: A critical target for cancer chemotherapy. *Clin. Colorectal Canc.* **1**, 220–229 (2002).
44. A. J. Patel, P. Varilly, D. Chandler, S. Garde, Quantifying density fluctuations in volumes of all shapes and sizes using indirect umbrella sampling. *J. Stat. Phys.* **145**, 265–275 (2011).
45. Z. Jiang, R. C. Remsing, N. B. Rego, A. J. Patel, Characterizing solvent density fluctuations in dynamical observation volumes. *J. Phys. Chem. B* **123**, 1650–1661 (2019).
46. N. B. Rego, E. Xi, A. J. Patel, Protein hydration waters are susceptible to unfavorable perturbations. *J. Am. Chem. Soc.* **141**, 2080–2086 (2019).
47. S. J. de Vries, A. M. J. J. Bonvin, How proteins get in touch: Interface prediction in the study of biomolecular complexes. *Curr. Protein Pept. Sci.* **9**, 394–406 (2008).
48. M. W. Turner, The role of mannose-binding lectin in health and disease. *Mol. Immunol.* **40**, 423–429 (2003).
49. T. C. Terwilliger, D. Eisenberg, The structure of melittin. I. Structure determination and partial refinement. *J. Biol. Chem.* **257**, 6010–6015 (1982).
50. L. L. Conte, C. Chothia, J. Janin, The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
51. S. Jones, J. M. Thornton, Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13–20 (1996).
52. I. M. Nooren, J. M. Thornton, Diversity of protein–protein interactions. *EMBO J.* **22**, 3486–3492 (2003).
53. U. M. Moll, O. Petrenko, The MDM2-p53 interaction. *Mol. Canc. Res.* **1**, 1001–1008 (2003).
54. J. M. Winget, T. Mayor, The diversity of ubiquitin recognition: Hot spots and varied specificity. *Mol. Cell* **38**, 627–635 (2010).
55. J. Janin, R. Bahadur, P. Chakrabarti, Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
56. C. Tse *et al.*, Exploring the free energy landscape and thermodynamics of protein-protein association. *Biophys. J.* **119**, 1226–1238 (2020).
57. I. A. Vakser, Protein-protein docking: From interaction to interactome. *Biophys. J.* **107**, 1785–1793 (2014).
58. E. S. C. Shih, M. J. Hwang, NPPD: A protein-protein docking scoring function based on dyadic differences in networks of hydrophobic and hydrophilic amino acid residues. *Biology* **4**, 282–297 (2015).
59. C. J. Fennell, K. A. Dill, Physical modeling of aqueous solvation. *J. Stat. Phys.* **145**, 209–226 (2011).

60. S. A. Paz, C. F. Abrams, Free energy and hidden barriers of the β-sheet structure of prion protein. *J. Chem. Theor. Comput.* **11**, 5024–5034 (2015).
61. P. Tiwary, J. Mondal, J. A. Morrone, B. Berne, Role of water and steric constraints in the kinetics of cavity–ligand unbinding. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12015–12019 (2015).
62. P. Tiwary, V. Limongelli, M. Salvalaglio, M. Parrinello, Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E386–E391 (2015).
63. E. Brini, D. Kozakov, K. A. Dill, Predicting protein dimer structures using MELD $\times$ MD. *J. Chem. Theor. Comput.* **15**, 3381–3389 (2019).
64. M. Kastelic, Y. V. Kalyuzhnyi, B. Hribar-Lee, K. A. Dill, V. Vlachy, Protein aggregation in salt solutions. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6766–6770 (2015).
65. K. P. Johnston *et al.*, Concentrated dispersions of equilibrium protein nanoclusters that reversibly dissociate into active monomers. *ACS Nano* **6**, 1357–1369 (2012).
66. N. J. Deng, P. Cieplak, Insights into affinity and specificity in the complexes of α-lytic protease and its inhibitor proteins: Binding free energy from molecular dynamics simulation. *Phys. Chem. Chem. Phys.* **11**, 4968–4981 (2009).
67. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
68. N. Chennamsetty, V. Voynov, V. Kayser, B. Helk, B. L. Trout, Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11937–11942 (2009).
69. C. C. Lee, J. M. Perchiacca, P. M. Tessier, Toward aggregation-resistant antibodies by design. *Trends Biotechnol.* **31**, 612–620 (2013).
70. M. Lund, L. Vrbka, P. Jungwirth, Specific ion binding to nonpolar surface patches of proteins. *J. Am. Chem. Soc.* **130**, 11582–11583 (2008).
71. K. Březina *et al.*, Can arginine inhibit insulin aggregation? A combined protein crystallography, capillary electrophoresis, and molecular simulation study. *J. Phys. Chem. B* **122**, 10069–10076 (2018).