

# Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects

Bryan A. Krantz<sup>1</sup>, Alok K. Srivastava<sup>2</sup>, Sehat Nauli<sup>3</sup>, David Baker<sup>3</sup>, Robert T. Sauer<sup>2</sup> and Tobin R. Sosnick<sup>1,4</sup>

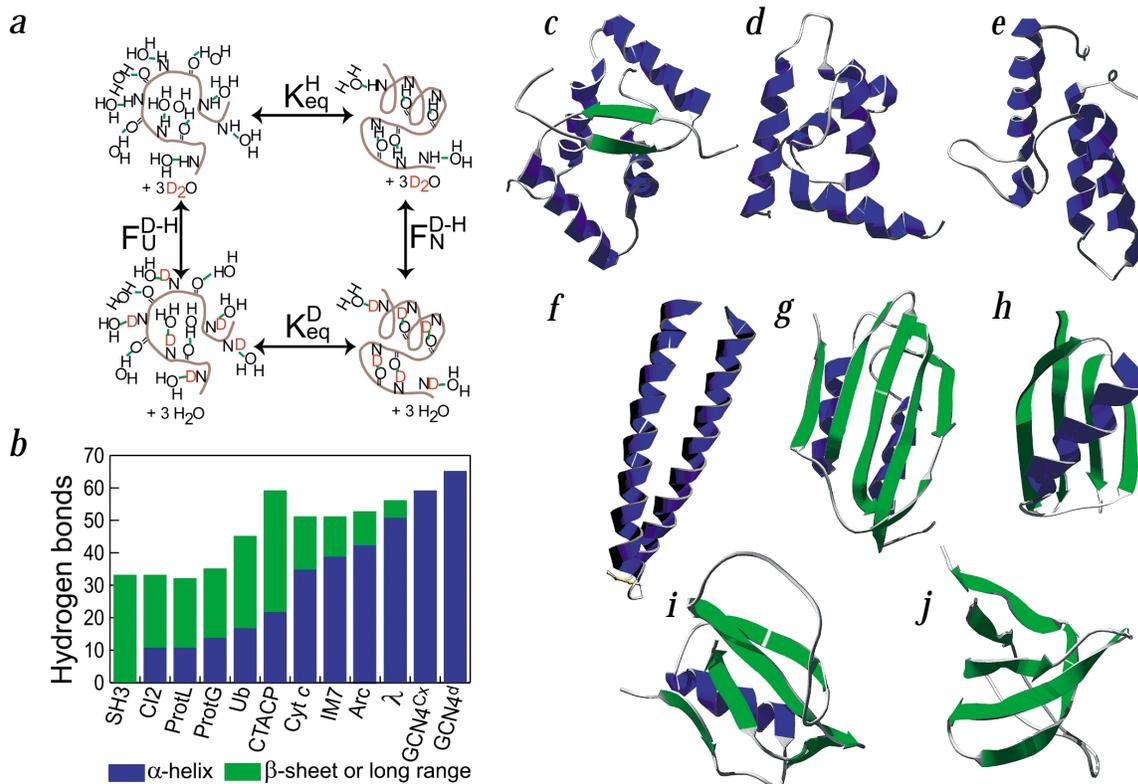
<sup>1</sup>Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58<sup>th</sup> Street, Chicago, Illinois 60637, USA. <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>Institute for Biophysical Dynamics, University of Chicago, 920 East 58<sup>th</sup> Street, Chicago, Illinois 60637, USA.

Published online: 29 April 2002, DOI: 10.1038/nsb794

**Through the development of a procedure to measure when hydrogen bonds form under two-state folding conditions,  $\alpha$ -helices have been determined to form proportionally to denaturant-sensitive surface area buried in the transition state. Previous experiments assessing H/D isotope effects are applied to various model proteins, including  $\lambda$  and Arc**

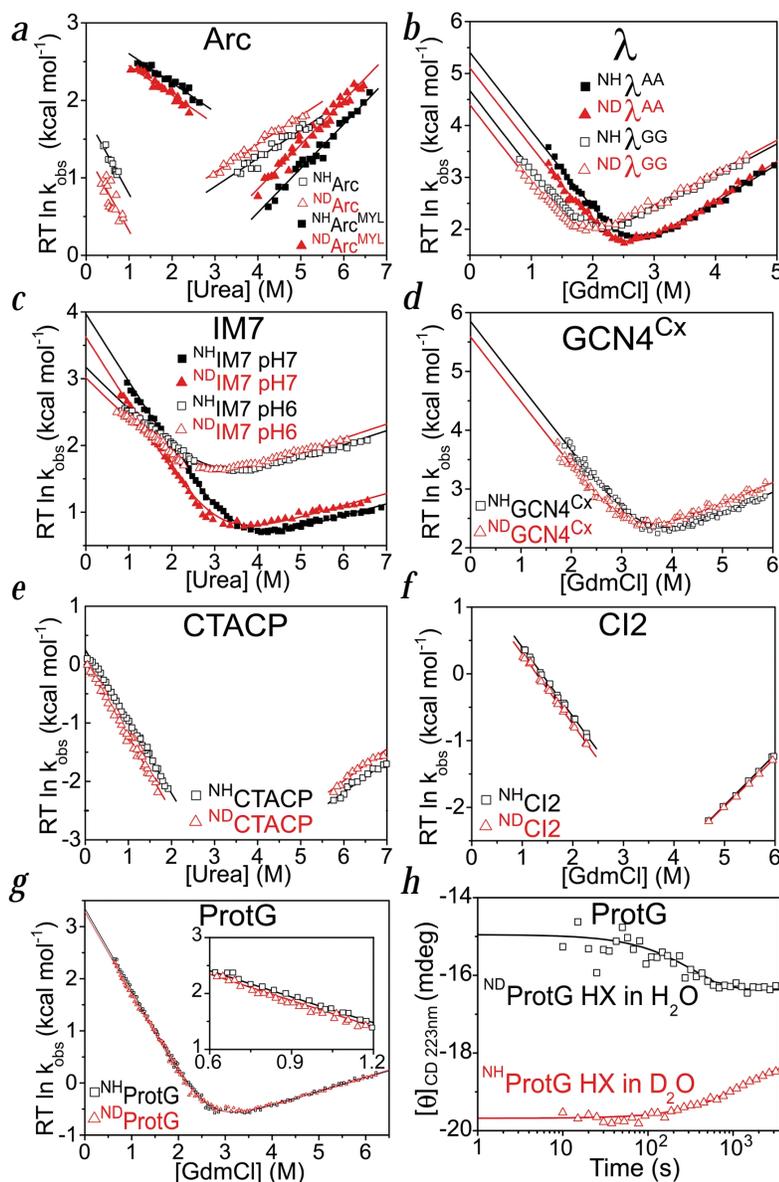
repressor variants, a coiled coil domain, cytochrome *c*, colicin immunity protein 7, proteins L and G, acylphosphatase, chymotrypsin inhibitor II and a Src SH3 domain. The change in free energy accompanied by backbone deuteration is highly correlated to secondary structure composition when hydrogen bonds are divided into two classes. The number of helical hydrogen bonds correlates with an average equilibrium isotope effect of  $8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ site}^{-1}$ . However,  $\beta$ -sheet and long-range hydrogen bonds have little isotope effect. The kinetic isotope effects support our hypothesis that, for helical proteins, hydrophobic association cannot be separated from helix formation in the transition state. Therefore, folding models that describe an incremental build-up of structure in which hydrophobic burial and hydrogen bond formation occur commensurately are more consistent with the data than are models that posit the extensive formation of one quantity before the other.

Biological reactions involving hydrogen (H) often can be dissected using deuterium (D) substitution<sup>1-3</sup>. The doubling of the mass of the hydrogen and the accompanying change in zero-point vibrational energy perturb chemical equilibria and kinetics, enabling the elucidation of the folding mechanism and transition state (TS) structure. In protein folding, kinetic amide isotope effects directly assay for the extent of hydrogen bond formation in the TS. Isotope effects overcome some of the limitations of mutational  $\phi$ -analysis, which assays for the degree of



**Fig. 1** Isotope effects and model systems. Amide isotope effects diagrammed as a thermodynamic cycle. **a**, Unfolded and folded protein models are on the left and right, respectively.  $K_{eq}^D$  and  $K_{eq}^H$  designate the equilibrium stability of protonated and deuterated protein, respectively. Global fractionation factors,  $F_U^{D-H}$  and  $F_N^{D-H}$ , are discussed elsewhere<sup>7</sup>. **b**, Bar graph showing hydrogen bond composition for small two-state model proteins<sup>7</sup>, where the number of helical hydrogen bonds are colored blue and the number of  $\beta$ -sheet hydrogen bonds (including a minor population of long-range hydrogen bonds) are green. Side chain-to-backbone hydrogen bonds are ignored because they are small in number, mostly solvent exposed and rapidly exchange compared with backbone amide hydrogen bonds<sup>35</sup>. Secondary structures are colored similarly in three-dimensional renderings of the following model proteins studied: **c**, Arc repressor (Arc); **d**,  $\lambda$ -repressor; **e**, colicin immunity protein (IM7); **f**, C-terminal crosslinked GCN4 (GCN4<sup>ca</sup>); **g**, common type acylphosphatase (CTACP); **h**, protein G (ProtG); **i**, chymotrypsin inhibitor II (Ci2); and **j**, Src SH3 domain (SH3). Protein L (ProtL) is not shown but topologically identical to ProtG. Renderings were created in the Swiss-Prot Protein Viewer (Glaxo Wellcome Experimental Research) and POV-Ray (<http://www.povray.org>).





**Fig. 2** Folding kinetic isotope effects. Kinetic activation energy plots for **a**, Arc and Arc<sup>MYL</sup> at pH 6 and 5, respectively; **b**,  $\lambda^{AA}$  and  $\lambda^{GG}$  at pH 7; **c**, IM7 at pH 6 and 7; **d**, GCN4<sup>Cx</sup> at pH 7; **e**, CTACP at pH 4.5; **f**, Cl2 at pH 4.5; and **g**, ProtG at pH 4.5. All measurements are conducted at 10 °C. Black squares and red triangles denote protonated and deuterated protein, respectively, for each chevron comparison. Standard stopped-flow methods (~1 ms dead time) are used; however, slower folding proteins (ProtG, CTACP and Cl2), which have smaller isotope effects, are also measured by rapid dilution with an automated titrator (dead time ~0.5 s). **h**, CD<sub>223nm</sub> equilibrium C<sub>m</sub> experiments for ProtG at pH 4.5 and 10 °C demonstrate that the 0.1 kcal mol<sup>-1</sup> isotope effect is equal and opposite when protonated protein is exchanged into D<sub>2</sub>O (red line with triangles) or deuterated protein is exchanged into H<sub>2</sub>O (black line with squares).

in folding activation energy relative to the change in equilibrium stability translated to a  $\phi_{I}^{D-H}$ -value, which reflected the global fraction of hydrogen bonds formed in the TS.

Here, we first examine whether isotope effect studies can provide information about hydrogen bond formation in both  $\alpha$ -helix and  $\beta$ -sheet secondary structures. Second, we ask whether hydrogen bond formation and surface area burial are concomitant processes in the TS. Third, if they are cooperative processes, what general aspects of folding pathways are related to surface burial? These issues are addressed using a variety of proteins studied under conditions where folding can be approximated by a two-state reaction (Fig. 1c–j). These systems sample the correlation between secondary structure formation and surface burial at values from 40 to 90% surface area burial, which is as wide of a sampling as can be expected for this parameter.

### Equilibrium isotope effects

Measurement of backbone amide isotope effects requires careful experimental design. Often the bulk solvent isotope effect<sup>5,10</sup> is of the same magnitude and can have the opposite sign as the amide isotope effect<sup>7</sup>. Control of this variable is obtained by conducting comparative

interaction of a mutated side chain and information about hydrogen bond formation is inferential.

However, only a few H/D amide isotope effect studies have been conducted under constant bulk solvent conditions<sup>4–7</sup>. Even though the innumerable noncovalent interactions in a protein should afford an isotope effect<sup>8</sup>, such investigations have been discouraged by the small isotope effects discovered in small molecule hydrogen bonds<sup>9</sup>. Furthermore, when a protein folding reaction is conducted in water, hydrogen bonds merely change partners; amides formerly bound to water in the unfolded state subsequently bind to carbonyl in the native state (Fig. 1a). This change is expected to diminish the relative importance of hydrogen bonds to protein stability as compared to other stabilizing forces, notably the hydrophobic effect.

With the goal of determining the degree of hydrogen bond formation in the TS, our group studied the folding amide isotope effect for the GCN4 coiled coil, cytochrome *c* (cyt *c*) and ubiquitin (Ub)<sup>7</sup>. The helical proteins were stabilized by ~0.5 kcal mol<sup>-1</sup> in the hydrogen form, whereas the mixed  $\alpha/\beta$  protein, Ub, was marginally stabilized by deuterium. The change

experiments in the same bulk solvent. Because backbone amides readily exchange with solvent, experiments are performed at low temperatures and pH (for example, 10 °C and pH 4.5–7), where hydrogen exchange (HX) is slowed<sup>11</sup>. This slowing enables accurate measurements of the stability of a deuterated protein in a protonated solvent before significant HX.

Equilibrium isotope effects,  $\Delta\Delta G^{D-H}$ , are measured by the C<sub>m</sub> experiment<sup>7</sup> in which deuterated protein sample is diluted 1:100 into a cuvette containing H<sub>2</sub>O and denaturant at a concentration where about half of the protein molecules are unfolded (Fig. 2h). Using far-UV CD, we monitor the change in stability that accompanies the replacement of the deuterated amides with hydrogens from the solvent, a process taking ~1 h. The initial and final CD readings are used to derive the equilibrium constant of the deuterated (K<sub>eq</sub><sup>D</sup>) and protonated (K<sub>eq</sub><sup>H</sup>) protein, respectively. Their ratio is used to calculate  $\Delta\Delta G^{D-H}$  (Eq. 1).

Kinetic studies (Fig. 2a–g) also assess equilibrium isotope effects,  $\Delta\Delta G^{D-H}$ , according to

$$\Delta\Delta G^{D-H} = -RT \ln (K_{eq}^D / K_{eq}^H) = -RT \ln ((k_f^D / k_u^D) / (k_f^H / k_u^H)) \quad (1)$$

Table 1 Equilibrium<sup>1</sup> and kinetic<sup>2</sup> amide isotope effects parameters

Protein	Denaturant / pH	$\Delta\Delta G^{D-H}$ (equil.)	$\Delta\Delta G^{D-H}$ (kinetics)	$\phi_r^{D-H}$	$m_f / m^o$
Arc <sup>3</sup>	Urea / 6	$-0.47 \pm 0.03^5$	$-0.65 \pm 0.08$	$0.72 \pm 0.08$	$0.76 \pm 0.02$
Arc <sup>MYL 3</sup>	Urea / 5	$-0.53 \pm 0.02^5$	$-0.46 \pm 0.06$	$0.34 \pm 0.07$	$0.40 \pm 0.01$
Cl2	GdmCl / 4.5	$-0.11 \pm 0.01$	$-0.16 \pm 0.04$	$0.88 \pm 0.28$	$0.60 \pm 0.01$
CTACP	GdmCl / 4.5	N.D.	$-0.33 \pm 0.04$	$0.59 \pm 0.15$	$0.72 \pm 0.02$
CTACP	Urea / 4.5	N.D.	$-0.40 \pm 0.03$	$0.50 \pm 0.08$	$0.70 \pm 0.04$
GCN4 <sup>Cx</sup>	GdmCl / 7	N.D.	$-0.43 \pm 0.04$	$0.61 \pm 0.03$	$0.75 \pm 0.02$
IM7	Urea / 6	N.D.	$-0.26 \pm 0.02$	$0.62 \pm 0.04$	$0.75 \pm 0.02$
IM7	Urea / 7	$-0.37 \pm 0.02$	$-0.50 \pm 0.03$	$0.69 \pm 0.02$	$0.85 \pm 0.01$
$\lambda^{AA 4}$	GdmCl / 4.5	$-0.39 \pm 0.01$	$-0.59 \pm 0.09$	$0.78 \pm 0.05$	$0.76 \pm 0.04$
$\lambda^{AA}$	GdmCl / 7	N.D.	$-0.36 \pm 0.02$	$0.84 \pm 0.04$	$0.67 \pm 0.01$
$\lambda^{AA}$	Urea / 7	N.D.	$-0.39 \pm 0.03$	$0.73 \pm 0.04$	$0.76 \pm 0.03$
$\lambda^{GG}$	GdmCl / 7	$-0.32 \pm 0.02^5$ ( $0.37 \pm 0.01$ ) <sup>5</sup>	$-0.34 \pm 0.02$	$0.80 \pm 0.03$	$0.71 \pm 0.02$
$\lambda^{GG}$	Urea / 7	N.D.	$-0.43 \pm 0.02$	$0.69 \pm 0.02$	$0.68 \pm 0.01$
ProtG	GdmCl / 4.5	$-0.10 \pm 0.02$ ( $0.10 \pm 0.01$ )	$-0.08 \pm 0.03$	$0.79 \pm 0.3$	$0.85 \pm 0.01$
ProtL	GdmCl / 4.5	$-0.04 \pm 0.02$ ( $0.06 \pm 0.01$ )	$-0.05 \pm 0.03$	$1.1 \pm 0.8$	$0.71 \pm 0.01$
SH3	GdmCl / 4.5	$-0.09 \pm 0.02$ ( $0.08 \pm 0.02$ )	$0.08 \pm 0.04$	$0.55 \pm 0.49$	$0.71 \pm 0.02$

<sup>1</sup> $C_m$  experiment, which measures  $\Delta\Delta G^{D-H}$ , upon exchange of deuterated protein in constant 99% bulk H<sub>2</sub>O solvent. Values in parentheses are for  $C_m$  experiments conducted in reverse — that is, constant 99% bulk D<sub>2</sub>O. The sign of  $\Delta\Delta G^{D-H}$  designates the difference in free energy between that starting and ending states. Units of  $\Delta\Delta G^{D-H}$  are kcal mol<sup>-1</sup>.

<sup>2</sup>Stopped-flow fluorescence kinetic measurements are conducted in constant ~90% H<sub>2</sub>O solvent. Titrator fluorescence kinetics experiments are conducted in constant ~99% H<sub>2</sub>O solvent.  $\Delta\Delta G^{D-H}$  and  $\phi_r^{D-H}$  are calculated from the middle of folding and unfolding arms of chevrons to minimize errors of extrapolation to 0 M denaturant.

<sup>3</sup>Free energy parameters of dimeric Arc<sup>MYL</sup> and wild type Arc are fit to data at experimental conditions of 9.6  $\mu$ M and 9.3  $\mu$ M monomer concentration, respectively.

<sup>4</sup>Stopped-flow CD at 222 nm probed the kinetics of <sup>ND</sup> $\lambda^{AA}$  at pH 4.5, because local HX events proximal to the chromophore potentially created fluorescence artifacts at pH 4.5. Fluorescence problems, however, were resolved when <sup>ND</sup> $\lambda^{AA}$  folding was measured at pH 7.

<sup>5</sup> $C_m$  experiment conducted at pH 4.5 instead of listed experimental pH.

Equality between the two independently derived values for  $\Delta\Delta G^{D-H}$  provides a stringent internal control (Table 1).

Hydrogen is often more stabilizing than deuterium, and the isotope effect is largest in helical proteins. These observations prompted analysis of the correlation between  $\Delta\Delta G^{D-H}$  and the number of helical hydrogen bonds in each protein (Fig. 3a). The correlation depends solely on the number of helical hydrogen bonds with a slope of  $-8.6 \pm 0.9$  cal mol<sup>-1</sup> site<sup>-1</sup>.  $\beta$ -sheet bonds seem to have little isotope effect, with a y-intercept of ~0 and random residuals. Isotope effects found for all  $\beta$ -sheet proteins, SH3 and CD2 (ref. 5), are negligible.

### Kinetic isotope effects

The kinetic isotope effect was measured with a four-syringe stopped-flow protocol and identical buffers to rigorously control for isotope and denaturant concentration<sup>7</sup>. For proteins with small isotope effects, <0.2 kcal mol<sup>-1</sup>, we developed a more accurate method that takes advantage of the fortuitously slow folding rates of these  $\alpha/\beta$ - and  $\beta$ -proteins: a titrator rapidly dilutes a deuterated or a protonated protein 100-fold into a cuvette under vigorous stirring. The experiment is highly controlled because the deuterated and protonated versions are alternatively injected into the same solution, thereby eliminating any difference in denaturant concentration, pH/pD and temperature (Fig. 2e–g).

Kinetic isotope effects were obtained using chevron analysis of the denaturant dependence of the folding and unfolding activation free energies<sup>12</sup>. The denaturant dependence of these quantities,  $\Delta G_f^\ddagger$  and  $\Delta G_u^\ddagger$ , is linear. The slope of the folding ( $m_f$ ) and unfolding ( $m_u$ ) arms relate to the amount of denaturant-accessible surface area buried in the pre- and post-folding TS, respectively. These slopes are the same for the protonated and deuterated versions of each protein. This feature enables the folding and unfolding kinetic isotope effects,  $\Delta\Delta G_f^{\ddagger D-H}$  and  $\Delta\Delta G_u^{\ddagger D-H}$ , respectively, to be determined from the average offset

of the respective chevron arms that occurs upon isotopic substitution. We find that the equilibrium isotope effect, determined from the kinetic isotope effect measurements (that is,  $\Delta\Delta G^{D-H} = \Delta\Delta G_f^{\ddagger D-H} - \Delta\Delta G_u^{\ddagger D-H}$ ) corroborates the value obtained from the equilibrium  $C_m$  measurements (Table 1; Fig. 3a).

### Analysis of kinetic isotope effects

The established methodology of mutational  $\phi$ -analysis<sup>12</sup> underlies the analysis of kinetic isotope effects. The effect of an amino acid substitution is quantified by the parameter  $\phi_r$ , given by the change in folding activation free energy,  $\Delta\Delta G_f^\ddagger$ , divided by the change in global stability,  $\Delta\Delta G_{eq}$ . A  $\phi_r$ -value is the degree to which the total energetic effect of the substitution is realized in the TS. Here, we perturb the stability by substituting all backbone amides with deuterons. The resulting isotopic  $\phi_r^{D-H}$ -value is

$$\phi_r^{D-H} = \Delta\Delta G_f^{\ddagger D-H} / \Delta\Delta G^{D-H} = RT \ln(k_f^D / k_f^H) / RT \ln(K_{eq}^D / K_{eq}^H) \quad (2)$$

The linear relationship between  $\Delta\Delta G^{D-H}$  and helical hydrogen bonds indicates that  $\phi_r^{D-H}$  provides information on  $\alpha$ -helical hydrogen bonds only. The  $\phi_r^{D-H}$  equals the fraction of hydrogen bonds formed in the TS, assuming that all helical hydrogen bonds contribute equally, they must either be formed or broken and they do not exist in a partially formed conformation. The all-or-none scenario is supported by our studies using engineered metal-binding sites<sup>13</sup> and mutational studies<sup>14</sup> of the coiled coil. In the TS, about half of the coiled coil molecule is structured while the remainder is largely unstructured. As  $\phi_r^{D-H} \sim 0.6$ , equating  $\phi_r^{D-H}$  to the fraction of native-like hydrogen bonds is more valid than the alternative — that is, all hydrogen bonds simultaneously are in an energetically strained or distorted geometry in the TS, each possessing a 60% isotope effect. Finally, because  $\beta$ -sheet hydrogen bonds are energetically insensitive to the isotopic substitution (Fig. 3a), the folding kinetics of a mixed



**Fig. 3** Correlating equilibrium isotope effects, kinetics and surface burial. **a**, Equilibrium isotope effects,  $\Delta\Delta G^{D-H}$ , are plotted for all systems studied against several helical hydrogen bonds. The slope,  $-8.6 \pm 0.9$  cal mol $^{-1}$  site $^{-1}$ , for the least squares linear fit (solid line) is the average isotope effect per helical hydrogen bond. The fit has an intercept of  $-13 \pm 40$  cal mol $^{-1}$  and R-value of 0.85. **b**, Denaturant-sensitive surface area burial ( $m_f/m^\circ$ ) is plotted versus folding hydrogen bond kinetic isotope effect ( $\phi_f^{D-H}$ ) for largely  $\alpha$ -helical proteins. A least-squares linear fit (solid line) has a slope of  $0.85 \pm 0.19$ , intercept of  $0.11 \pm 0.13$  and a R-value of 0.78, where dotted lines designate 97% upper and lower confidence bands. Figure includes previous data<sup>7</sup> and a version of cyt c mutated at the two peripheral His residues H26N and H33N (B.A.K., H. Rumbley, L. Mayne, S.W. Englander and T.R.S., unpublished data).

$\alpha/\beta$  protein may be analyzed with isotope effects, but only in terms of its  $\alpha$ -helical hydrogen bonds.

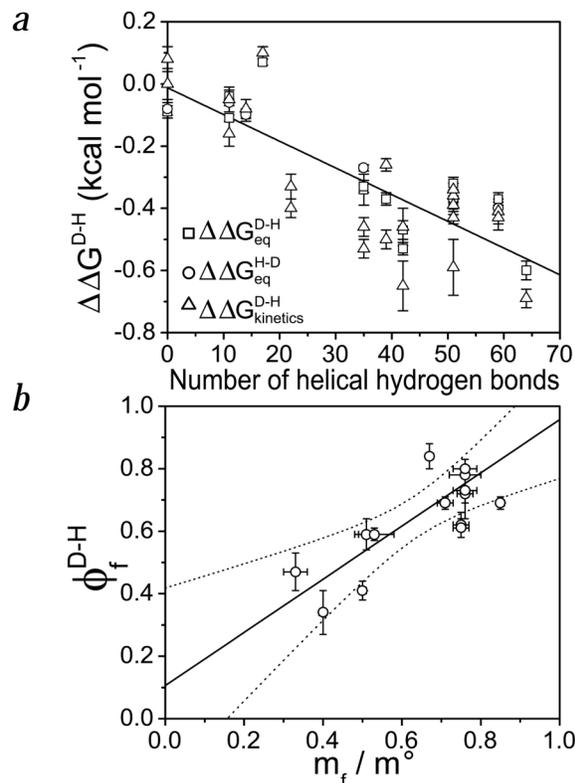
### Model proteins

**Arc repressor.** In the small, dimeric transcription factor Arc, ~75% of its backbone hydrogen bonds are helical, whereas 8–15% are involved in the two  $\beta$ -strands (Fig. 1). Waldburger *et al.*<sup>15</sup> dissected the importance of an internal salt bridge and hydrogen bond formed by a triad of side chains. When replaced with hydrophobic side chains (R31M, E36Y and R40L), this Arc<sup>MYL</sup> variant folds 10- to 1,000-fold faster than wild type, depending upon denaturant concentration. Furthermore,  $m_f/m^\circ$  shifted from 0.75 to 0.4 in Arc<sup>MYL</sup>. The  $\phi_f^{D-H}$ -values for Arc and Arc<sup>MYL</sup> are  $0.72 \pm 0.08$  and  $0.34 \pm 0.07$ , respectively (Fig. 2a), providing a strong correlation with  $m_f/m^\circ$  (Table 1).

Srivastava and Sauer<sup>16</sup> investigated helix formation in the TS of Arc with Ala $\rightarrow$ Gly substitutions. They found a high average  $\phi_f$ -value ( $0.82 \pm 0.56$ ) for helix B and a low average  $\phi_f$ -value ( $0.26 \pm 0.22$ ) for helix A. In combination with the isotope effect data that indicates that 72% of the helical hydrogen bonds are formed in the TS, we surmise that most of helix B, but only half of helix A, forms in the TS. By extension, Arc<sup>MYL</sup> probably has only a portion of helix B formed in the TS.

Interestingly, Arc<sup>MYL</sup> forms fewer hydrogen bonds in the TS than wild type but folds faster<sup>15</sup>, indicating that more extensive helical structure formation in the TS does not necessarily result in faster folding. Furthermore, the near diffusion-limited folding rate of Arc<sup>MYL</sup>,  $2 \times 10^8$  M $^{-1}$  s $^{-1}$  at 25 °C, indicates that the major fraction of chain encounters are productive. Assuming that helix B is formed in the TS, as indicated by the mutational and isotope effect data, folding is two to three orders of magnitude too fast to be contingent on transient pre-formed helix B. In such a diffusion/collision scenario,  $k_f$  is only  $4 \times 10^5$  M $^{-1}$  s $^{-1}$ ; that is, the product of the calculated 0.05% probability of both monomers being helical<sup>17</sup> and the diffusion-limited encounter rate is  $\sim 10^9$  M $^{-1}$  s $^{-1}$ , assuming 100% encounter success frequency.

**$\lambda$ -repressor.** We choose to study two forms of the five-helix bundle protein,  $\lambda$ -repressor: wild type ( $\lambda^{GG}$ ) and the G46A/G48A mutant ( $\lambda^{AA}$ ). The double Gly mutation slightly reduces the extent to which the 51  $\alpha$ -helical hydrogen bonds form in the TS. Generally, for both mutants,  $\phi_f^{D-H}$  and  $m_f/m^\circ$  are well correlated, with ~70–80% of the  $\alpha$ -helical hydrogen bonds formed (although the ratio for  $\lambda^{AA}$  in guanidinium,  $1.25 \pm 0.06$ , the largest deviation we observe, suggests a larger percentage of hydrogen bonds is formed in this system). Consistently, Oas *et al.*<sup>18</sup> observed high mutational (Ala $\rightarrow$ Gly)  $\phi_f$ -values in helices 1, 4 and 5, which account for 67% of the helical content.



**Colicin immunity protein IM7.** The colicin immunity protein IM7 is a four-helix bundle. At pH 6 and 7, the folding behavior is consistent with a two-state model — for example, chevron arms are linear with denaturant concentration. In addition, all the formation of hydrogen bonds can be accounted for in the observed reaction, because the equilibrium  $\Delta\Delta G^{D-H}$  is recapitulated from the kinetic measurements (Table 1). Finally, the folding kinetic isotope effect remains constant, which is inconsistent with the accumulation of a highly helical intermediate at lower denaturant concentrations. For IM7, we conclude that ~70% of the helical hydrogen bonds form at the TS, in agreement with recent mutational studies by Radford and coworkers<sup>19</sup>.

**Acylphosphatase CTACP.** CTACP is composed of five  $\beta$ -strands and two  $\alpha$ -helices. Because the  $\beta$ -sheet hydrogen bonds are insensitive to isotopic substitution, the  $\phi_f^{D-H}$ -value is 50%, indicating that half of the  $\alpha$ -helical bonds are formed in the TS. Using mutational  $\phi_f$ -analysis, Taddei *et al.*<sup>20</sup> concluded that only helix 2 is formed in the TS, in agreement with the  $\phi_f^{D-H}$ -value. The kinetically important Pro 54 ( $\phi_f = 0.98$ ), located in the initiating turn for helix 2, may guide formation of the  $\beta$ -strands and provide other crucial hydrophobic context for the helix in the TS.

**Chymotrypsin inhibitor CI2.** For this small  $\alpha/\beta$  protein, Fersht *et al.*<sup>21</sup> found that mutational  $\phi_f$ -values are the most significant in the single helix (11 hydrogen bonds), with an average value of  $0.44 \pm 0.27$ . However, because few of the helical residues have significantly high  $\phi_f$ -values ( $>0.7$ ), they concluded that the helix is either partly formed or possessed an extended geometry in the TS. Although the equilibrium isotope effect is small, 110–160 cal mol $^{-1}$ , it is consistent with the 11 helical hydrogen bonds. The  $\phi_f^{D-H}$ -value is near unity, indicating that most of the helical hydrogen bonds are formed in the TS.

**Protein G, GB1 domain.** Protein G has 35 backbone amides involved in hydrogen bonds, 14 of which are involved in a single  $\alpha$ -helix. Baker *et al.*<sup>22</sup> determined that the average (non-negative)  $\phi_T$ -value in the  $\alpha$ -helix is  $0.25 \pm 0.13$ . The  $\phi_T^{D-H}$  is  $0.79 \pm 0.30$ , suggesting that a larger proportion of the  $\alpha$ -helix is present in the TS. This inference, however, should be viewed with caution because the magnitude of the isotope effect is small.

The apparent discrepancy in hydrogen bond content between the mutational and isotope studies for CI2 and protein G may provide an example in which either side chain interactions, which can be tertiary in nature, do not exclusively reflect the backbone hydrogen bond network or the  $\phi_T$ -values under-report the structure<sup>23</sup>. Lower mutational  $\phi_T$ -values in the helix with a higher value for the isotope effect may indicate that hydrogen bond formation occurs with fewer native-like side chain contacts. This uncertainty can be resolved using engineered metal-binding sites and  $\Psi$ -analysis, a methodology that can identify the fraction of pathways having a given helix formed in the TS<sup>13</sup>.

### General folding models

For largely  $\alpha$ -helical proteins at the rate-limiting step, the degree of helix formation, given by  $\phi_T^{D-H}$ , is proportional to the degree of surface area buried, as quantified by  $m_T/m^\circ$  (0.78 R-value, Fig. 3*b*). At the TS, secondary and tertiary structures form concomitantly in predominantly  $\alpha$ -helical proteins. We infer that this relationship is continually maintained throughout the folding reaction, as observed in folding simulations of a three-helix bundle<sup>24</sup>. However, for the  $\beta$ -sheet and mixed  $\alpha/\beta$  systems, the correlation between  $\phi_T^{D-H}$  and  $m_T/m^\circ$  is poor (0.24 R-value). Proteins with extensive  $\beta$ -sheet networks would not be expected to show such a correlation because of the absence of an isotope effect in  $\beta$ -sheets.

The correlation of  $\phi_T^{D-H}$  and  $m_T/m^\circ$  values allows us to evaluate the applicability of various folding models for  $\alpha$ -helical proteins. If extensive hydrophobic collapse occurred before helix formation, the percentage of hydrogen bonds formed should have been less than the amount of surface area buried. Furthermore, such a collapse with marginal backbone desolvation is unstable<sup>25</sup>, because the failure to form each compensating backbone hydrogen bond costs at least several kcal mol<sup>-1</sup>. Additionally, the two-state folding behavior observed in nearly all proteins under  $\sim 100$  amino acids indicates that stable, non-specific collapse does not occur prior to the TS. These observations argue against a nonspecific hydrophobic collapse model.

Likewise, if extensive hydrogen bond formation preceded tertiary interactions, the percentage of hydrogen bond formed should have been larger than we observed. Also, folding contingent on preformed helical structure is contrary to results with Arc<sup>15</sup>, as discussed above, and studies of the dimeric GCN4 coiled coil<sup>14</sup> and a variant with enhanced tertiary interactions<sup>26</sup>. Additionally, folding rates for Arc and GCN4, like most helical proteins, are faster at higher temperatures, contrary to the expectation for a simple folding scenario in which rates are limited by precollision helical structure — that is, as temperature increases, helical content decreases more significantly than the diffusion rate increases.

The interpretation that the TS is a critically distorted version of an essentially native-like structure with a desolvated core (for example, the 'dry molten-globule' model) also does not apply to helical proteins, because many hydrogen bonds are not formed and significant surface remains exposed at the TS.

The present data support the hypothesis that protein folding is an incremental build up of structure, where hydrophobic burial

and hydrogen bond formation occur commensurately<sup>7</sup>, rather than a model where extensive formation of either one of these quantities occurs before the other. The TS has extensive regions of presumably near-native structure, whereas other regions remain largely unfolded and exposed to solvent, as observed in the unfolding intermediates of cyt *c*<sup>27</sup> and ribonuclease H<sup>28</sup> identified using native state HX methods.

Previously, we proposed that the critical element of the TS is the formation of the overall chain topology, which is established by pinning the chain through the interaction of several apolar side chains to adequately define the native topology<sup>29,30</sup>. Consistently, a strong correlation was observed for proteins between the folding speed and contact order<sup>31</sup>. At the time of our original proposal, we focused on TS topology driven by hydrophobic association and not on secondary structure formation. The present isotope study revises this view and establishes that hydrophobic association cannot be separated from helix formation. Evidently, when the hydrophobic side chains of an amphipathic  $\alpha$ -helix coalesce, water — the major competitors of backbone hydrogen bonds — is expelled, which drives bonding between the amide and carbonyl groups. Once these hydrogen bonds are formed, the solvent-exposed hydrogen bonds on the opposite side of the now nucleated helices form. In this view, helix formation provides an efficient means to satisfy the hydrogen-bonding requirements of hydrophobically collapsed polypeptide. These principles should apply to hydrogen bond formation in  $\beta$ -sheets as well, even though their edges can be solvated.

Isolated helix formation does not drive the folding process. Rather, folding occurs as a result of hydrophobic association. The energetic benefit of hydrophobic association offsets the otherwise unfavorable formation of isolated helical structure. This situation is analogous to helix formation inside a membrane. Here, the equilibrium constant for helix formation increases greatly because the alternative, exposed backbone groups are now highly unfavorable. Thus, the observed correlation that hydrogen bonds form in proportion to surface area burial may simply reveal that the exclusion of water molecules necessitates the formation of hydrogen bonds rather than the intrinsic strength of helix formation driving its own formation.

During the folding process, the polypeptide creates its own desolvated environment<sup>6,32</sup>. Transition states generally bury a significant amount of surface area, as indicated by the general observation that  $m_T$ -values are minimally  $\sim 1$  and 0.5 units in guanidinium-HCl and urea, respectively. Hence, there may be a size threshold for backbone desolvation and nucleation. Which regions become desolvated and result in structure formation in the TS depend on the sequence and structure of the protein. Without the formation of correct topology, apolar interactions will disassemble and re-assemble while searching for a collapse-competent nucleus. This topological requirement, which depends upon the structure of a protein, may explain the wide diversity in the degree of denaturant sensitive surface area burial and extent of hydrogen bond formation, from  $\sim 40\%$  to  $\sim 85\%$ . Likewise, a distinct topological, rather than stability-based requirement, rationalizes the observation that the amount of surface burial in the TS is typically independent of whether the protein is under stabilizing and destabilizing conditions.

### Conclusions

Amide isotope effects provide a unique, powerful method to assess hydrogen bond formation in  $\alpha$ -helices with minimal perturbation to the system. We find that helical hydrogen bond formation and surface area burial occur in a concerted

fashion. Models that heavily weigh either of these two quantities seem contradictory to our results. Backbone desolvation accompanying hydrophobic association results in helix formation, because it is an efficient means to satisfy the hydrogen bonding requirements. Therefore, in combination with site-specific data, these insights may best direct theoretical simulations.

## Methods

**Proteins.** Expression plasmids were constructed using pRSET B (Invitrogen) with subcloned open reading frames (Operon technologies) for IM7 (ref. 19),  $\lambda^{AA}$  (ref. 18), Cl2 (ref. 21), CTACP<sup>20</sup> and SH3 (ref. 33) proteins. Protein L, protein G, Arc and Arc<sup>MYL</sup> were obtained as described<sup>15,34</sup>. Wild type  $\lambda^{GG}$  was created using Stratagene Quick Change mutagenesis kit. GCN4<sup>Cx</sup> is a C-terminally crosslinked version of GCN4-p1' (ref. 13). Dissolving lyophilized protein in deuterated guanidinium-HCl produced deuterated proteins.

**Equilibrium.** CD was used to monitor changes in stability at wavelengths from 222–228 nm ( $\pm 5$  nm) using a Jasco 715 at concentrations of 2–10  $\mu$ M.

**Kinetics.** Experiments used a Biologic stopped-flow<sup>7</sup> at concentrations of 0.1–10  $\mu$ M.

**Analysis.** Kinetic data were analyzed using chevron analysis<sup>12</sup> in which the free energy,  $\Delta G$ , and activation free energy for folding,  $\Delta G^{\ddagger}_f$ , and unfolding,  $\Delta G^{\ddagger}_u$ , are linearly dependent on denaturant. The dependence on denaturant, or m-values, assesses the degree of surface area burial during folding. When kinetic folding reactions are two-state, equilibrium values are calculated by  $\Delta G = \Delta G^{\ddagger}_f - \Delta G^{\ddagger}_u$  and  $m^{\circ} = m_u - m_f$ .

## Acknowledgments.

We thank A. Fernandez, X. Fang, S.W. Englander, N. Kallenbach, C. Brooks, R.S. Berry, T. Pan and our group members for numerous enlightening discussions. This work was supported by grants from the NIH and The Packard Foundation Interdisciplinary Science Program (T.R.S., P. Thiyagarajan, S. Berry, D. Lynn and S. Meredith).

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to T.R.S. *email:* trsosnic@midway.uchicago.edu

Received 21 December, 2001; accepted 26 March, 2002.

- Northrop, D.B. *Annu. Rev. Biochem.* **50**, 103–131 (1981).
- Schowen, K.B. & Schowen, R.L. *Methods Enzymol.* **87**, 551–606 (1982).
- Cleland, W.W. *Methods Enzymol.* **249**, 341–373 (1995).
- Itzhaki, L.S. & Evans, P.A. *Protein Sci.* **5**, 140–146 (1996).
- Parker, M.J. & Clarke, A.R. *Biochemistry* **36**, 5786–5794 (1997).
- Kentsis, A. & Sosnick, T.R. *Biochemistry* **37**, 14613–14622 (1998).
- Krantz, B.A., Moran, L.B., Kentsis, A. & Sosnick, T.R. *Nature Struct. Biol.* **7**, 62–71 (2000).
- Wade, D. *Chem. Biol. Interact.* **117**, 191–217 (1999).
- Jancso, G.V.H. & Alexander, W. *Chem. Rev.* **74**, 689–719 (1974).
- Makhatadze, G.I., Clore, G.M. & Gronenborn, A.M. *Nature Struct. Biol.* **2**, 852–855 (1995).
- Connelly, G.P., Bai, Y., Jeng, M.-F., Mayne, L. & Englander, S.W. *Proteins* **17**, 87–92 (1993).
- Matthews, C.R. *Methods Enzymol.* **154**, 498–511 (1987).
- Krantz, B.A. & Sosnick, T.R. *Nature Struct. Biol.* **8**, 1042–1047 (2001).
- Moran, L.B., Schneider, J.P., Kentsis, A., Reddy, G.A. & Sosnick, T.R. *Proc. Natl. Acad. Sci. USA* **96**, 10699–10704 (1999).
- Waldburger, C.D., Jonsson, T. & Sauer, R.T. *Proc. Natl. Acad. Sci. USA* **93**, 2629–2634 (1996).
- Srivastava, A.K. & Sauer, R.T. *Biochemistry* **39**, 8308–8314 (2000).
- Munoz, V. & Serrano, L. *Biopolymers* **41**, 495–509 (1997).
- Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L. & Oas, T.G. *Nature Struct. Biol.* **4**, 305–310 (1997).
- Capaldi, A.P., Kleanthous, C. & Radford, S.E. *Nature Struct. Biol.* **9**, 209–216 (2002).
- Taddei, N. *et al. J. Mol. Biol.* **300**, 633–647 (2000).
- Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. *J. Mol. Biol.* **254**, 260–288 (1995).
- McCallister, E.L., Alm, E. & Baker, D. *Nature Struct. Biol.* **7**, 669–673 (2000).
- Bulaj, G. & Goldenberg, D.P. *Nature Struct. Biol.* **8**, 326–330 (2001).
- Guo, Z., Brooks, C.L. III & Boczek, E.M. *Proc. Natl. Acad. Sci. USA* **94**, 10161–10166 (1997).
- Honig, B. & Yang, A.S. *Adv. Protein. Chem.* **46**, 27–58 (1995).
- Durr, E., Jelesarov, I. & Bosshard, H.R. *Biochemistry* **38**, 870–880 (1999).
- Bai, Y., Sosnick, T.R., Mayne, L. & Englander, S.W. *Science* **269**, 192–197 (1995).
- Chamberlain, A.K., Handel, T.M. & Marqusee, S. *Nature Struct. Biol.* **3**, 782–787 (1996).
- Sosnick, T.R., Mayne, L., Hiller, R. & Englander, S.W. *Peptide and Protein Folding Workshop* (ed. DeGrado, W.F.) 52–80 (International Business Communications, Philadelphia; 1995).
- Sosnick, T.R., Mayne, L. & Englander, S.W. *Proteins* **24**, 413–426 (1996).
- Plaxco, K.W., Simons, K.T. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
- Fernández, A. *J. Chem. Phys.* **114**, 2489–2502 (2001).
- Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L. & Serrano, L. *Biochemistry* **33**, 2142–2150 (1994).
- Nauli, S., Kuhlman, B. & Baker, D. *Nature Struct. Biol.* **8**, 602–605 (2001).
- Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. *Proteins* **17**, 75–86 (1993).