research papers

On the determinants of amide backbone exchange in proteins: a neutron crystallographic comparative study

Brad C. Bennett,[a] Anna S. Gardberg,[b] Matthew D. Blair[b] and Chris G. Dealwis[a,b]*

[a]Department of Pharmacology, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4965, USA, and [b]Biochemistry, Cellular and Molecular Biology Department, University of Tennessee-Knoxville, Knoxville, TN 37996-0001, USA

Correspondence e-mail:
chris.dealwis@case.edu

The hydrogen/deuterium-exchange (HDX) method, coupled with neutron diffraction, is a powerful probe for investigating molecular dynamics. In the present report, general determinants of HDX are proposed based on 12 deposited neutron protein structures. The parameters that correlate best with HDX are the depth within the protein structure of the amide N atom and the secondary-structure type. Both the B factor of the amide N atom and the ratio $B/\langle B \rangle$ correlate moderately. However, solvent accessibility only correlates strongly for one molecule and hydrogen-bonding distance correlates for two molecules with respect to amide HDX. In addition to the relatively small number of neutron structures available, the limitations to this type of analysis, namely resolution, data completeness and the data-to-parameter ratio, are discussed briefly. A global analysis of HDX was performed to overcome some of these obstacles, damping the effects of outliers and the extreme variation of the data sets arising from resolution limitations. From this, amide depth and hydrogen-bonding distance to the amide (a measure of interaction strength) show strong global correlation with HDX. For some structures, the constituents of the hydrophobic protein core could be identifed based on contiguous regions that are resistant to exchange and have significant depth. These may, in fact, constitute minimal folding domains.

1. Introduction

Although neutron macromolecular crystallography has been available as a tool for structural biologists since 1969 (Schoenborn, 1969), it has experienced a resurgence within this decade, with the number of protein neutron structures deposited in the Protein Data Bank (PDB; Berman *et al.*, 2002) more than doubling. A major reason for this is the implementation of three diffraction instruments at three different neutron sources in the last 10 y: the Laue Diffracto-meter (LADI) at the Insitut Laue-Langevin (ILL; Cipriani *et al.*, 1996; Niimura *et al.*, 1997) in 1997, BIX-3 at Japan Atomic Energy Research Institute (JAERI) in 1999 (Ostermann *et al.*, 2002) and the Protein Crystallography Station (PCS) at the Los Alamos Neutron Scattering Center (LANSCE) in 2002 (Langan *et al.*, 2004). The advent of spallation neutron sources allows data to be collected efficiently with a low background using the time-of-flight (TOF) Laue method. The sole instrument for single-crystal biological neutron diffraction (ND) at a spallation source is currently in operation at the aforementioned PCS at LANSCE (Langan *et al.*, 2004). Recent proof-of-principle studies from PCS include xylose isomerase (XI; Katz *et al.*, 2006; Hanson *et al.*, 2004) and dihydrofolate reductase (DHFR; Bennett *et al.*, 2006). In the

Table 1
Characteristics of reported neutron protein structures and/or diffraction data sets.

Entries in bold indicate the proteins for which HDX analysis was performed. NP, data not provided.

Structure (PDB code, if any)	MW (kDa)	Crystal volume† (mm³)	Soak pH (pD)	Soak time (d)	Soak temperature (K)	Collection time (d)	Unit-cell volume‡ (Å³)	Space group	Resolution (Å)	Solvent content (%)
1. DHFR (2inq)§	18	0.7	7.9	30	277	22.5	5.18×10^5	$P6_1$	2.2	52
2. Aldose reductase D¶	35.9	0.15	5	D_2O	298	NP	1.62×10^5	$P2_1$	2.2	46
3. β-Trypsin (1ntp)	23.3	5	7	365	293	NP	2.17×10^5	$P2_12_12_1$	1.8	45
4. Concanavalin A (1c57)	25.6	4.8	6.5	21	NP	9.6	4.94×10^5	$I222$	2.4	48
5. Concanavalin A (1xqn)	25	5.6	6.5	D_2O	298	17.0	4.73×10^5	$I222$	2.5	48
6. DsrD (1wq2)§	8.8	1.7	5.3	120	293	70.0	1.83×10^5	$P2_12_12_1$	2.4	53
7. Endothiapepsin (1gkt)	33.8	3.5	4.5	70	298	69.0	1.39×10^5	$P2_1$	2.1	39
8. Insulin (3ins)	11.6	3.5	6.5	135	NP	NP	2.00×10^5	$H3$	2.2	36
9. Lysozyme (1lzn)	14.3	2	4.7	180	298	14.0	2.62×10^4	$P1$	1.7	32
10. Lysozyme (1io5)	14.3	6	7	NP	291	NP	2.42×10^5	$P4_32_12$	2.0	42
11. Myoglobin (1l2k)	17.2	6.25	6.8	3650	298	24.0	6.69×10^4	$P2_1$	1.5	37
12. Myoglobin–CO (2mb5)	17.2	24	5.7	Months	NP	NP	6.71×10^4	$P2_1$	1.8	37
13. Myoglobin-D (1cq2)¶	17.2	2.5	6.5	244	298	NP	6.67×10^4	$P2_1$	2.0	37
14. RNase A (5rsa)	13.7	30	5.3	183	NP	60.0	5.94×10^4	$P2_1$	2.0	43
15. RNase A (6rsa)	13.7	19	6.8	90	NP	NP	5.99×10^4	$P2_1$	2.0	44
16. Rubredoxin (1vcx)	5.9	5	8	NP	298	35.0	5.36×10^4	$P2_12_12_1$	1.5	46
17. Rubredoxin-mut (1iu6)	5.9	4	8.5	NP	298	30.0	5.31×10^4	$P2_12_12_1$	1.6	46
18. Subtilisin	27	2.7	6.1	30	NP	305.0	2.15×10^5	$P2_12_12_1$	2.0	38
19. Trp repressor	13	1	5	30	293	13.4	9.34×10^4	$P2_12_12$	2.1	32
20. Xylose isomerase (2gve)	43.3	8	8	30	293	19.0	9.63×10^5	$I222$	1.8	56
21. BPTI (5pti)	6.5	8	8.2	60	NP	NP	5.01×10^4	$P2_12_12_1$	1.8	36
22. Rasburicase	135	1.8	8.5	D_2O	298	34.5	4.07×10^5	$I222$	2.1	58.7
23. DFPase	35	0.43	6.5	7	NP	37	3.11×10^5	$P2_12_12_1$	2.2	45.5
24. Amicyanin	11.5	2.6	4.8	D_2O	NP	21	4.31×10^4	$P2_1$	1.9	34.4
25. Crambin	4.9	1.4	7.4	Weeks	NP	NP	1.72×10^4	$P2_1$	1.5	32
26. PYP	14	0.79	7	30	293	13.75	1.58×10^5	$P6_3$	2.5	35
27. Insulin (2efa)	11.6	2.7	6.6	NP	298	18.8	4.94×10^5	$I2_13$	2.7	34
28. Deoxyhemoglobin (2dxm)	64.6	20	6.3	NP	NP	18	2.89×10^5	$P2_1$	1.8	45.1
Average	23.10	5.82	6.52	327.81	295	48.81	207234.52		1.97	41.95
Median	23.32	6.02	6.46	347.67	298.00	50.57	194280.29		1.96	42.15

Structure (PDB code, if any)	Backbone HDX†† (%)	R_{free} (%)	R_{work} (%)	R_{merge} (%)	Temperature (K)	Neutron source	Completeness (%)	Reference
1. DHFR (2inq)§	A, 65; B, 57	23.3	18.2	7	293	LANSCE	79.7	Bennett et al. (2006)
2. Aldose reductase D¶	NP	32	26	22.8	293	ILL	73.5	Blakeley et al. (2008)
3. β-Trypsin (1ntp)	76	NA	19	NP	293	HFBR	NP	Kossiakoff (1982)
4. Concanavalin A (1c57)	NP	30.1	27	22.2	293	ILL	89	Deacon et al. (1997)
5. Concanavalin A (1xqn)	NP	32	26.6	14.3	15	ILL	76	Blakeley et al. (2004)
6. DsrD (1wq2)§	A, 45; B, 62	29	23	14.3	293	JAERI	92.5	Chatake et al. (2003)
7. Endothiapepsin (1gkt)	49	27.4	23.5	7.5	293	ILL	84.5	Coates et al. (2001)
8. Insulin (3ins)	100	NA	19.1	10.4	NP	NBS	60	Wlodawer et al. (1989)
9. Lysozyme (1lzn)	70	22	20	14.6	NP	ILL	83	Bon et al. (1999)
10. Lysozyme (1io5)	100	32	21	16.4	291	ILL	81.4	Niimura et al. (1997)
11. Myoglobin (1l2k)	89	30.6	28.3	10.3	298	JAERI	87.9	Ostermann et al. (2002)
12. Myoglobin–CO (2mb5)	91	NP	11.5	3.8	NP	HFBR	NP	Cheng & Schoenborn (1990)
13. Myoglobin-D (1cq2)¶	NP	25	16	10.6	298	HFBR	96	Shu et al. (2000)
14. RNase A (5rsa)	75	NA	18.3	5.5-7.1	NP	NBS	51	Wlodawer et al. (1986)
15. RNase A (6rsa)	83	NA	20.7	6.3	NP	NBS	73	Borah et al. (1985)
16. Rubredoxin (1vcx)	89	21.7	18.6	9.6	298	JAERI	81.9	Kurihara et al. (2004)
17. Rubredoxin-mut (1iu6)	72	22.8	20.1	9.1	299	JAERI	87.5	Li et al. (2004)
18. Subtilisin	NP	NA	19.2	10	NP	HFBR	68	Kossiakoff et al. (1991)
19. Trp repressor	65	31.3	30.4‡‡	11.4	293	ILL	72.6	Lawson & Chin (2003)
20. Xylose isomerase (2gve)	75	31.9	27.1	26.2	295	LANSCE	78	Katz et al. (2006)
21. BPTI (5pti)	79	NA	19.7	7.6	NP	NBS	NP	Wlodawer et al. (1984)
22. Rasburicase	NP	NP	NP	15.1	293	ILL	64.1	Budayova-Spano et al. (2006)
23. DFPase	NP	NP	NP	19.9	293	LANSCE	81.8	Blum et al. (2007)
24. Amicyanin	NP	NP	NP	11.6	293	LANSCE	68.2	Sukumar et al. (2005)
25. Crambin	NP	NA	14.4	NP	293	HFBR	73	Teeter & Kossiakoff (1984)
26. PYP	47.5	27.3	26.2	31	293	LANSCE	88.8	Fisher et al. (2007)
27. Insulin (2efa)	NP	29.1	21.6	14.6	293	JAERI	95.7	Ishikawa et al. (2008)
28. Deoxyhemoglobin (2dxm)	NP	NP	NP	21.9	293	LANSCE	76.1	Kovalevsky et al. (2008)
Average	79.50	27.94	21.15	12.77	278.72		77.60	
Median	79.50	28.29	20.96	13.05	277.88		77.28	

† No correction for crystal morphology. ‡ Corrected for Bravais lattice. § Two molecules per asymmetric unit. ¶ Perdeuterated. ††Values computed from PDB file. ‡‡2003 model.

past, a major deterrent against ND has been the requirement for crystal volumes in excess of 1 mm^3. However, for DHFR and diisopropyl fluorophosphatase (DFPase), ND data have been collected at PCS (Blum *et al.*, 2007) using crystal volumes of 0.3 and 0.43 mm^3, respectively. A plot of crystal sample volumes against the respective unit-cell volumes of proteins for which ND data have been collected (listed in Table 1) is shown in Fig. 1.

The importance of ND in the study of biological systems is its utility in the accurate determination of H-atom positions, which is vital information that other direct structural techniques (such as X-ray crystallography and NMR) typically cannot provide. H atoms invariably play a role in a multitude of processes from acid–base-catalyzed enzymatic reactions to noncovalent interactions that drive and strengthen receptor-ligand binding. Whereas X-ray scattering intensity depends on an atom's electronic content (atomic Z number), neutron scattering lengths are independent of increasing Z. Hydrogen can be visualized by neutrons because it has a large total scattering cross-section (82.03 barns). However, hydrogen's negative coherent neutron scattering length (−3.74 fm) and large incoherent scattering cross-section (80.27 barns) contributes to the background. Fortunately, its heavier stable isotope deuterium (D) scatters neutrons with a 40-fold reduction in incoherence (2.05 barns) while possessing a positive scattering length (+6.67 fm) very similar to that of carbon (+6.65 fm) (Sears, 1992). Hence, replacing H with D dramatically improves the signal-to-noise-ratio.

A prerequisite for nearly every macromolecular ND study performed to date is to exchange H for D in the protein crystal. This is performed in one of four ways: (i) by using D$_2$O-containing reservoir buffers at crystallization setup (Sukumar *et al.*, 2005), (ii) by soaking or vapor-exchanging the crystal with deuterated mother liquor (Bennett *et al.*, 2005, 2006), (iii) by using perdeuterated protein in the crystallization setup with deuterated mother liquor and reservoir buffer (Hazemann *et al.*, 2005) or (iv) by using perdeuterated protein in the crystallization setup with hydrogenated buffers and then soaking or vapor-exchanging the fully grown crystals with deuterated buffer (Shu *et al.*, 2000). The first two methods are the most common by far for introducing D into protein crystals. Labile H positions (those in hydroxyl, carboxylate, amide and amine groups) can be exchanged for D using either of these methods. Perdeuteration involves synthesizing proteins from deuterated amino acids, where H at chemically non-exchangeable positions (*i.e.* H bound to an aliphatic carbon) is replaced by D. This dramatically enhances coherent scattering from the crystal, thus enhancing signal to noise.

Nuclear density peaks for D atoms are readily distinguishable given data at 2.5 Å (or better) resolution (Myles, 2006). The ability to visualize these atoms allows the definition of the protonation states of key residues [such as in trypsin (Kossiakoff & Spencer, 1981), endothiapepsin (Coates *et al.*, 2001), XI (Katz *et al.*, 2006) and DHFR (Bennett *et al.*, 2006)] and to provide a more robust solvent-structure model [for example, in lysozyme (Bon *et al.*, 1999) and concanavalin A (Habash *et al.*, 2000; Blakeley *et al.*, 2004)] because now all

three solvent atoms can be modeled. Lastly, and most important for the present analysis, because crystals are grown in the presence of or exchanged with D$_2$O-containing buffer and H/D can be visualized in medium- to high-resolution nuclear density maps, the degree of H/D exchange at each backbone amide position may then be derived.

In a series of elegant experiments half a century ago, Linderstrøm-Lang and colleagues at the Carlsberg laboratories demonstrated that hydrogen on backbone amide groups exchanged with deuterated solvent and used this method to test Pauling's idea that backbone hydrogen bonding was the major stabilizer of secondary-structural elements in proteins (Hvidt & Linderstrom-Lang, 1954; Englander *et al.*, 1997). Depending on the pH (or pD), amide backbone hydrogen/deuterium exchange (HDX) is caused by transient interaction of the amide group with solvent D$_3$O$^+$ (acid-catalyzed) or OD$^-$ (base-catalyzed) ions. At physiological pH, the exchange rate of amide groups within model peptides is somewhat slow and is on a suitable time scale for ensemble rate measurements (~0.5 s^{-1}) using practical instrumentation (Creighton, 1993). Thus, HDX is a powerful tool for understanding protein structure and dynamics (Englander *et al.*, 1996; Clarke & Itzhaki, 1998; Raschke & Marqusee, 1998) while being a considerably mild probe relative to other 'labels' commonly used (such as larger stable or radioactive isotopes or fluorescent dyes). This is because the deuteron itself is relatively small and can penetrate into the depths of a macromolecule, allowing exchange to reflect both residue-level and global (molecule-level) dynamics (Kossiakoff, 1982).
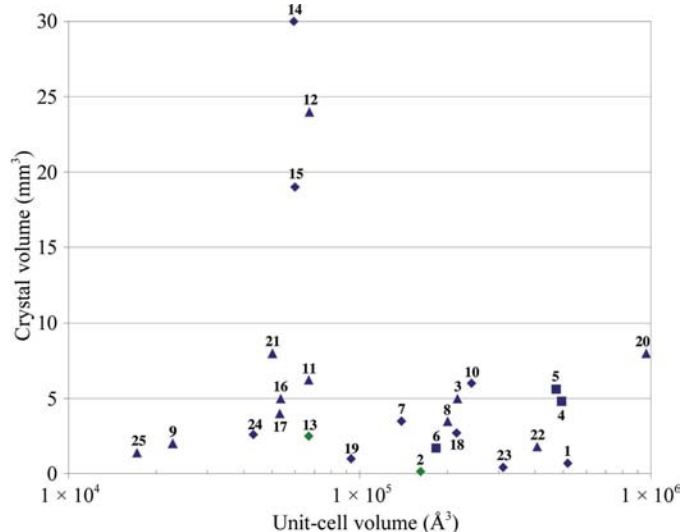

Figure 1
A plot of protein crystal sample volume against the respective volume of the crystal unit cell for all reported neutron diffraction data sets. For clarity, data points (individual proteins) are labeled with the number assigned to them in Table 1. Crystal sample volumes have not been corrected for the external morphology of the crystal (*i.e.* whether or not the crystal is a perfect cube), while unit-cell volumes have been corrected for the respective Bravais lattice type. Additionally, the data points are categorized by the resolution limit of the data set: triangles indicate 1.5–1.8 Å resolution, diamonds indicate 1.9–2.2 Å resolution and squares indicate 2.3–2.5 Å resolution. Green data points indicate perdeuterated protein crystals.

The protective effect from exchange imparted by hydrogen bonding in secondary structure is well documented (Fersht, 1998). Hydrogen bonding and the burial of amide groups in the core of the protein have recently been used as key parameters to devise a molecular-dynamics-based calculated protection factor for any desired position in a protein structure. The calculated factors from simulations for α-lactalbumin agree well (correlation coefficient of >0.8) with experimental data; this shows that secondary structure imparts significant protection from exchange and reveals populations of molecules undergoing structural fluctuations that may be necessary (indeed, may be the basis) for exchange (Vendruscolo et al., 2003). The mechanism by which HDX occurs for an amide within the protein core and/or within a highly structured region is still not well understood (Englander et al., 1997). There are two hypotheses for the process: one invokes global, subglobal or local unfolding of the polypeptide chain (Englander, 2000), of which the latter two can be involved in protein 'breathing' motions (Englander et al., 1980), while the other concerns the diffusion of solvent to access core regions of the protein (Kossiakoff, 1982). The former postulates that an amide can exchange owing to local melting or even global unfolding of the protein structure (Kossiakoff, 1982; Loh et al., 1993; Bai et al., 1994), exposing the amide so as to allow access to bulk (deuterated) solvent. The latter hypothesis explains HDX at core residues by a solvent-permeation mechanism, whereby atomic vibrations and channels within the molecule allow D_2O molecules to intrude into the protein interior (Creighton, 1993). Both of these factors (unfolding and permeation) are likely to play a role in determining the HDX propensity of a given amide. However, the opposite is also true: amides within core regions that have resisted exchange have done so because of the impenetrability of solvent or the inability (energetically) to unfold core structural elements or a combination of both.

The above mechanisms describe the overall HDX (or HDX protection) in a protein. However, what factors explain the observed exchange differences between backbone amide positions on the polypeptide on a residue-by-residue level? Are there measurable parameters that can help one determine whether a backbone amide is likely to undergo HDX or be protected from it? Analysis and quantification of HDX is now often performed by mass spectrometry (monitoring an increase in mass in intact protein or hydrolyzed peptides; Maier & Deinzer, 2005; Chalmers et al., 2006; Yamamoto et al., 2004) or NMR (often monitoring the disappearance of the correlated ^{15}N–^1H cross-peaks in the HSQC spectra as the NH group exchanges to ND; Wagner & Wuthrich, 1982; Polshakov et al., 2006). With these and other solution techniques, one can measure overall HDX as well as time-resolved HDX by quantifying exchange rates for groups of amides over a time course. Time-dependent HDX rates at different pH levels were measured for two proteins of similar mass, insulin and bovine pancreatic trypsin inhibitor (BPTI), by Hvidt & Pedersen (1974). They found that at physiological or moderately basic pH, insulin undergoes near-complete exchange much faster than BPTI. However, even for the structurally rigid BPTI molecule less than 10% of exchangeable H atoms remained after 10 h (Hvidt & Pedersen, 1974). Li & Woodward (1999) grouped the amide H atoms slowest to exchange into a 'slow-exchange core' and proposed that the slow-exchange core consisted of the regions that organize early in protein folding (Li & Woodward, 1999). The advantage of using ND over other methodologies as a tool for analyzing HDX is straightforward. The ability to directly visualize D atoms in nuclear density maps, to model them into the structure and to refine their occupancies provides, depending on the resolution of the data (Habash et al., 1997), a robust and accurate method for determining the HDX pattern of a protein.

Limitations on the occupancy refinement and its accuracy include the resolution of the data, the completeness of the data and the data-to-parameter ratio. Additionally, the number of reported neutron structures (28) restricts the analysis, simply because the sample size is relatively small, especially when one considers the structures for which detailed HDX information is available (only ten of the 28). The problem with attempting to refine D occupancies at resolutions at or around 2.75 Å has been demonstrated by the Helliwell laboratory for the neutron structure refinement of concanavalin A (Habash et al., 1997). For this reason, at resolutions lower than 2.5 Å deuterium-occupancy refinement should only be performed with much caution and apprehension of the output values. Once deuterium-occupancy refinement has been performed, however, the subsequent analysis allows one to understand better the dynamics of the protein at the whole-molecule level as well as at the residue level. The earliest work that showed the power of coupling ND with HDX to probe protein dynamics was performed by Wlodawer on RNase A (Wlodawer & Sjolin, 1982); subsequently, Kossiakoff revealed the HDX pattern for trypsin (Kossiakoff, 1982). However, to our knowledge there is no comprehensive study of the HDX patterns and properties for the neutron protein structures deposited within the PDB. Therefore, using nearly all the deposited structures for which backbone-exchange information is available, we have investigated the relationship between HDX and protection with secondary structure, atomic depth, B factors, absolute solvent accessibility (ASA) and hydrogen-bonding environment. The present work hopes to identify any molecular determinants of HDX that are conserved across proteins. An additional goal is to allow a researcher armed with an X-ray or NMR structure but with neither ND nor HDX data to better understand the dynamics of the protein under study.

2. Experimental methods

2.1. Definitions

In this work, which is an analysis of the backbone-exchange behavior of proteins as revealed from their respective neutron crystal structures, we are interested specifically in the backbone amide N atoms and their properties. As such, the atom-depth, B-factor and solvent-accessibility values analyzed and

discussed within this report are solely for the backbone amide N atoms. Atom depth is defined here as the distance in angstroms between a given amide N atom and its nearest solvent-accessible atom. Solvent accessibilities reported here are the solvent-accessible surface areas (in Å^2) of backbone amide N atoms using a water probe of radius 1.4 Å.

For all of our HDX correlation graphs, the y axis represents the percentage of exchanged amides. We did this because the number of available data points in some of the bins was significantly smaller relative to the other bins. Additionally, we acknowledge that the categorization of the coefficients of determination (the R^2 values from linear regression analysis) from strong to weak is only semi-quantitative. We consider an R^2 of >0.7 as strong and of 0.5–0.64 as moderate; any value falling below 0.5 is considered as weak or poor correlation.

2.2. Protein selection

We could only select protein structures for this analysis if refined occupancies (from 0.0 to 1.0) for D atoms bound to the backbone amide were provided in the deposited PDB file. Additionally, when multiple coordinate sets existed for a protein, we selected a representative for the analysis; indeed, this was the case for myoglobin and lysozyme. An exception to this is RNase A; we analyzed both of the deposited structures (5rsa and 6rsa) in order to perform a comparison of the apo form of the protein with its ligand-bound complex. The following 12 structures (with PDB codes in parentheses) were included in the HDX correlation analysis: DHFR monomers A and B (2inq), DsrD monomers A and B (1wq2), endothiapepsin (1gkt), RNase A (5rsa and 6rsa), rubredoxin (1vcx), xylose isomerase (2gve), BPTI (5pti), myoglobin (1l2k) and lysozyme (1lzn). Because two monomers exist in the asymmetric unit for DHFR and DsrD, we treat them here as two separate neutron structures.

It should be noted that as a measure of HDX we concern ourselves only with refined occupancy values of D atoms bound to backbone amide N atoms. We are not attempting to measure or extrapolate exchange rates. Additionally, we extracted the refined occupancy values directly from the coordinates as deposited in the PDB. We consider a refined deuterium occupancy of >0.2 to mean that there has been sufficient deuterium exchange at the particular amide backbone position to be detected. Thus, the amide to which the deuterium is bound must come into contact with D_2O at least transiently. Below this value, we consider that deuterium substitution is very weak. Therefore, from a statistical standpoint, HDX at these positions would in actuality be highly unlikely.

We then set backbone exchange as an admittedly simplified and binary relationship: occupancies of >0.2 indicate amide positions with a propensity for deuterium substitution (exchanged) and occupancies of <0.2 indicate positions with a propensity to be protected from deuterium substitution (not exchanged). We also chose this strategy for delineating exchange in order to prevent over-interpretation of the occupancy data. No additional refinement was performed on

any of the neutron structures. Additionally, no errors or standard deviations of the measurement of the occupancy values were available or are reported here. The inability to estimate errors of the occupancy values is a limitation of an in-depth HDX study at present. As more proteins are subjected to neutron diffraction and more of these structures are refined jointly against a combination of isomorphous X-ray and neutron data, it may be that estimates of the deviations in experimentally determined values will become more suitable to calculate. For the structures selected for HDX analysis, we emphasize that these are derived from quite different data resolutions, ranging from 1.5 to 2.4 Å. Thus, we cannot rule out that some of the variability observed in our analysis could be attributed to this inherent difference.

2.3. Data extraction and analysis

Data were extracted from the PDB files and combined for analysis with the grep and awk commands and *via* Perl scripts. Where deuterium occupancies (Docc) were reported, they were used. Where hydrogen occupancies (Hocc) were reported, the formula Docc = 1 − Hocc was used to calculate Docc. In the case of a mixed file with some H or D occupancies reported and not others, residues lacking a reported amide H/D-occupancy value were omitted from the analysis, except in the case of endothiapepsin, for which such residues (other than proline) were assigned Docc = 0 (L. Coates, personal communication). All proline residues were omitted.

2.4. Secondary structure

The *DSSP* algorithm (Kabsch & Sander, 1983) was used to assign secondary structure to each residue. The categories used were α-helix, isolated β-bridge (β-strand with a length of 1), extended β-strand, 3_{10}-helix, 5-helix (π-helix), hydrogen-bonded turn and bend (region of high curvature). Residues remaining unassigned by *DSSP* are in loops or of irregular structure not involved in hydrogen bonding; this type is often denoted 'random coil', but Kabsch and Sander (creators of the *DSSP* algorithm, see Kabsch & Sander, 1983) eschew this terminology. Instead, amides that are not assigned to a particular defined secondary-structure type are termed to reside in unstructured or randomly structured regions. While crystallographically ordered, these residues lack hydrogen-bonding structure, so this class of residue was adopted as the equivalent of 'unstructured peptide' (Molday *et al.*, 1972) for computation of relative protection factors.

2.5. Atom depth

The *DPX* algorithm to calculate atom depth has been described in Pintar & Pongor (2003). Surface residues are defined to have zero depth, while the depth for interior residues is defined as the distance to the closest atom that possesses a solvent-accessible surface [the *DPX* program runs *NACCESS* (Hubbard & Thornton, 1996) to determine solvent accessibilities]. We used *DPX* to specifically determine the depth of the amide N atom. Each amide N atom was classified by its depth and exchange status (whether its bound H atom
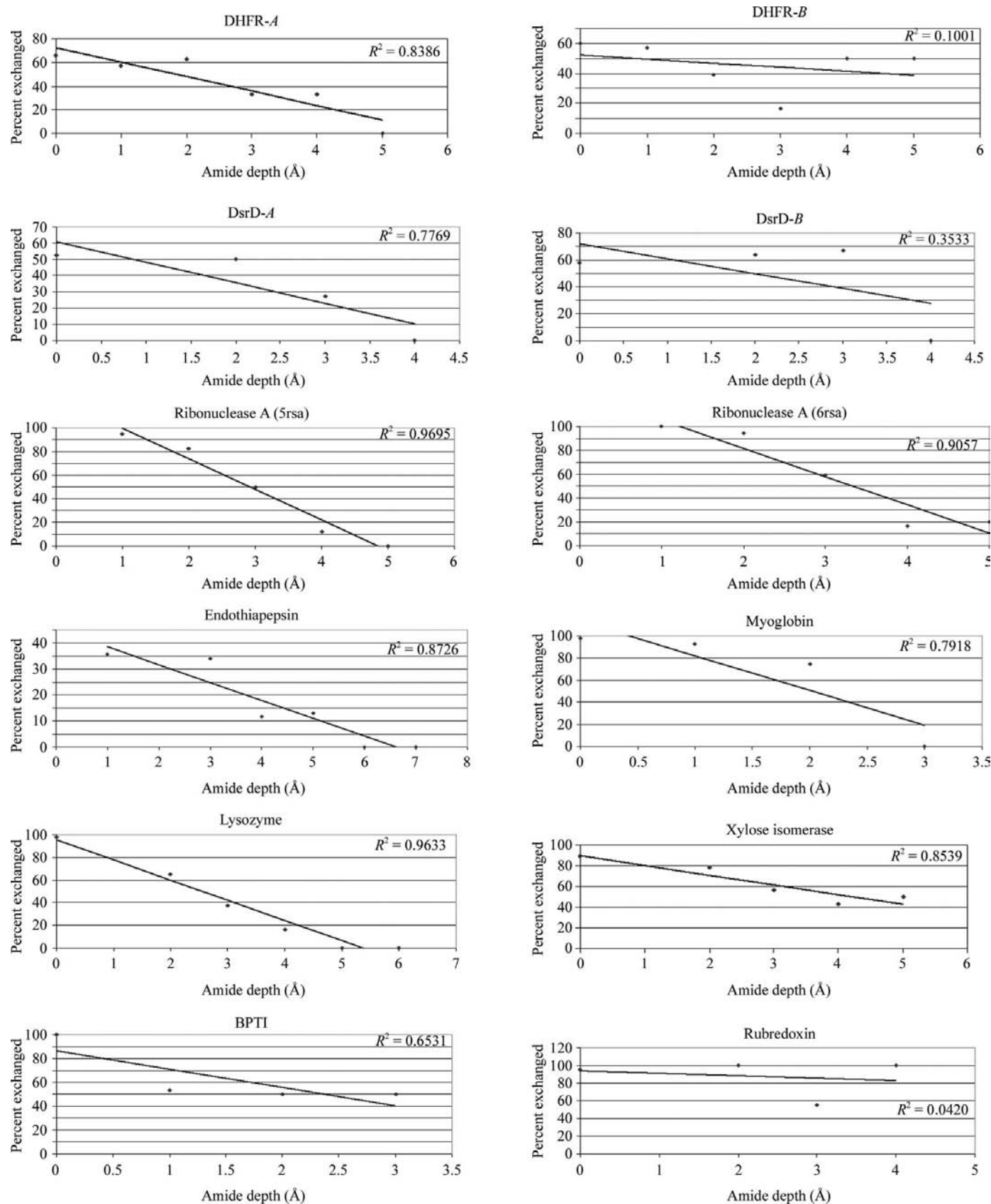
Figure 2
Correlation plots of HDX and amide depth. Individual HDX–depth correlation plots are shown for all the protein molecules included in the analysis. For nine of the 12 proteins tested, there is a strong correlation ($R^2 > 0.65$) between atom depth and the propensity for the amide to undergo exchange. Depth was measured by *DPX*. The *y* axis for all plots (and also those in Figs. 3–6) is the percentage of fully exchanged residues.
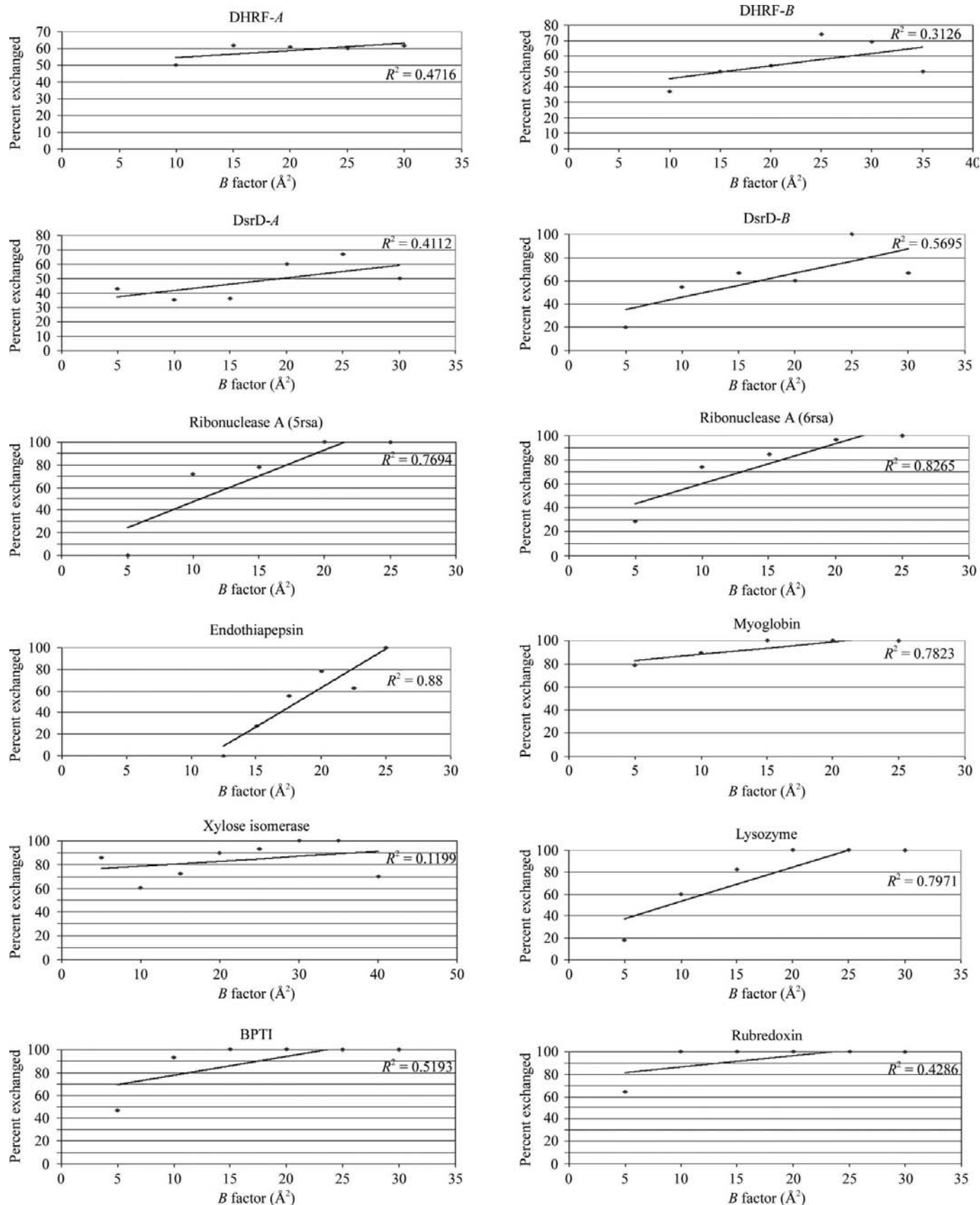
Figure 3
Correlation plots of HDX and amide isotropic B factor. Individual HDX–temperature factor correlation plots are shown for all the protein molecules included in the analysis. This is the next best correlative to HDX (besides secondary structure), with five of the 12 proteins tested showing strong correlation.
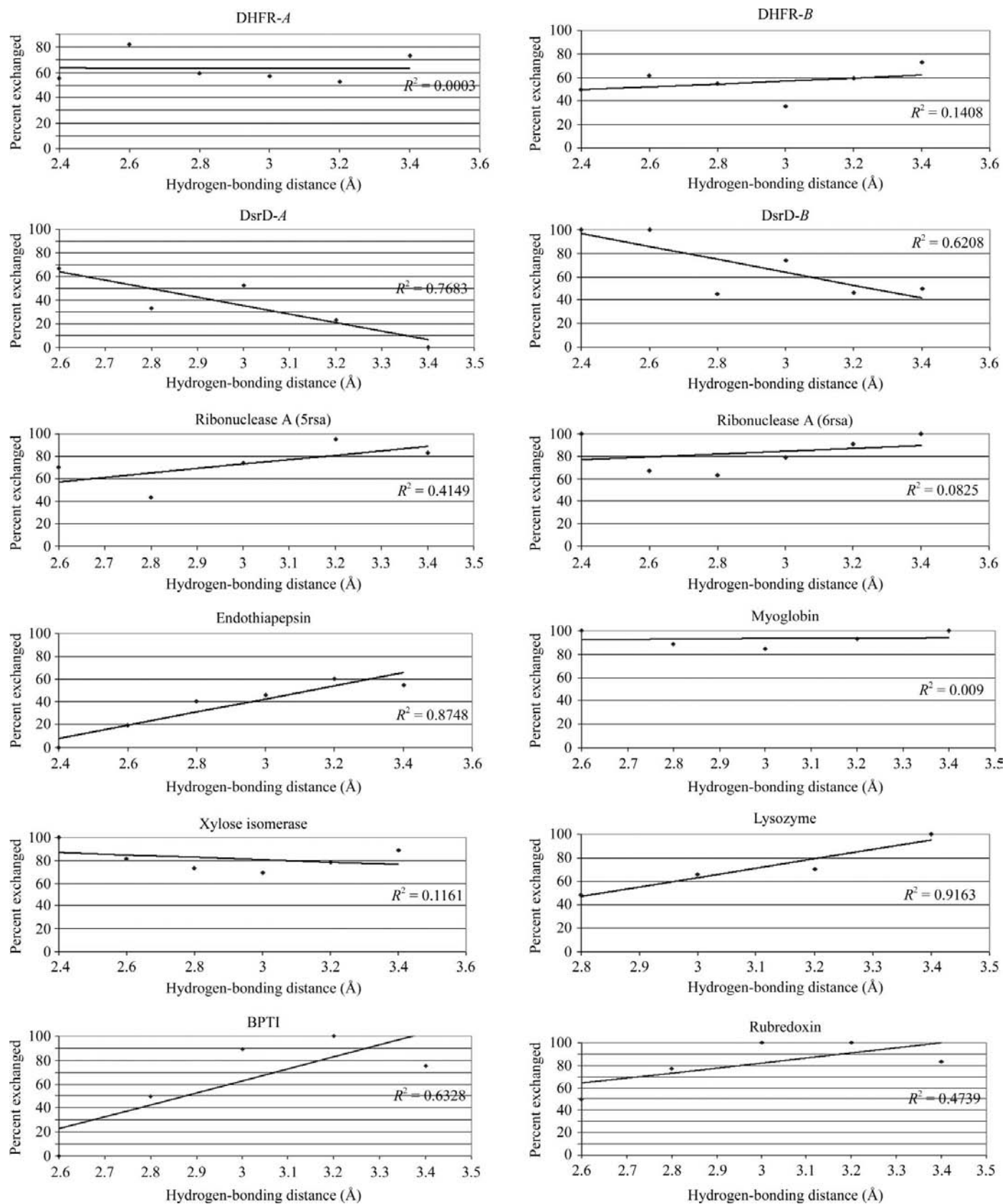
Figure 4
Correlation plots of HDX and amide hydrogen-bonding distance. Individual HDX–amide hydrogen-bonding distance correlation plots are shown for all the protein molecules included in the analysis. Two of the proteins show strong interdependence between HDX and amide hydrogen-bonding distance.
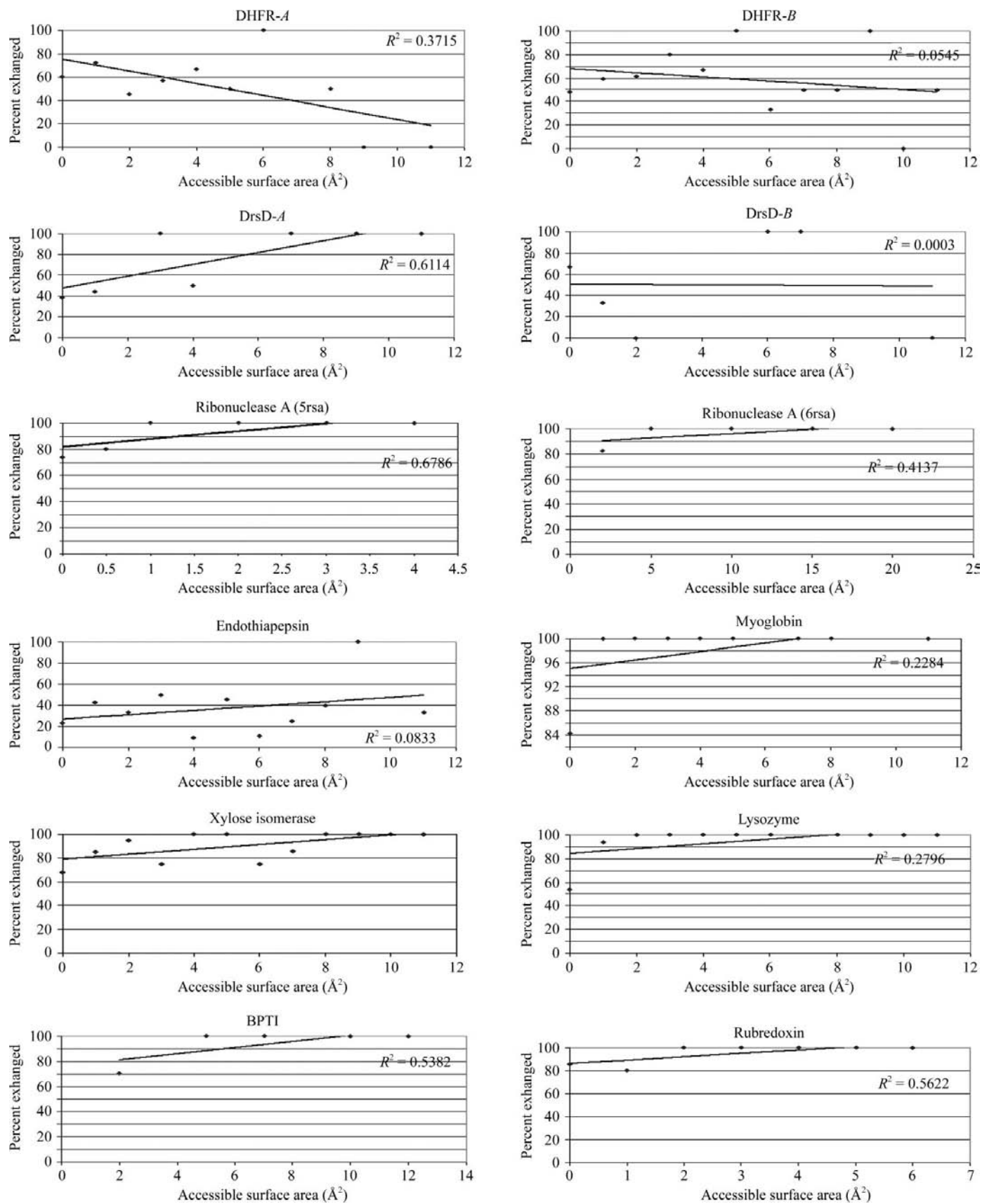
Figure 5
Correlation plots of HDX and amide solvent accessibility. Individual HDX–amide accessibility correlation plots are shown for all the protein molecules included in the analysis. RNase A is the only protein which shows a strong correlation between HDX propensity and the solvent-accessible surface of the amide.

had exchanged for D). Gradations of 1 Å depth, normally sectioned into six bins (0, 1, 2, 3, 4 and >5 Å), were used to bin the data for most of the structures. For example, the 0 bin is actually representative of the depth range 0–1 Å, the 1 bin represents depths that fall between 1 and 2 Å and so on. For some proteins, depths were binned slightly differently, with gradations at 1.25 or 0.5 Å. For each depth, the fraction of residues at that depth that had exchanged was calculated and a coefficient of determination was derived for each crystal structure.

2.6. Isotropic B factors

Refined B values for all amide N atoms were extracted from PDB coordinate files. Individual amide N atoms were classified by B factor and exchange status. For most proteins, the percentage of fully exchanged and non-exchanged residues in six bins of B-factor range (0–5.00, 5.01–10.00, 10.01–15.00, 15.01–20.00, 20.01–25.00 and >26.00 Å^2) were used to calculate the coefficients of determination. For endothiapepsin, gradations of 2.5 Å^2 were used.

In an attempt to normalize B factors because of resolution and refinement-program variability, a calculation of the ratio of the amide B factor to the arithmetic mean amide B factor ($B/\langle B \rangle$) in a given structure was performed. This provides an idea of the deviation of B factors from the average B value. $B/\langle B \rangle$ values were binned in the following divisions: 0.1–0.5, 0.6–1.0, 1.1–1.5, 1.6–2.0 and >2.0. The percentage of fully exchanged residues was plotted against the various $B/\langle B \rangle$ bins; trend lines were fitted and coefficients of determination were calculated.

2.7. Hydrogen-bonding distance

Hydrogen-bonding contacts made to the backbone amide N atoms of all 12 neutron structures were determined using the *CCP*4 program *NCONT*. Specifically, we defined the distance range for a hydrogen bond as 2.4–3.5 Å with all possible donor/acceptor atom types (such as O, S, P and N). Distances were rounded to the nearest tenth of an angstrom. In this calculation, only one donor and acceptor for the amide was included. We binned these contacts by their hydrogen-bonding distance to the amide using narrow distance ranges (2.4–2.6, 2.7–2.9, 3.0–3.2 and 3.3–3.5 Å). We categorized the amides by their HDX and generated plots of the percentage of exchanged amides *versus* distance(s) to their respective hydrogen-bonding partner(s).

2.8. Solvent accessibility

The *AREAIMOL* program (Collaborative Computational Project, Number 4, 1994) was used to compute the absolute solvent-accessible surface area (ASA) for every amide N atom. A solvent-sphere probe radius of 1.4 Å was used (Lee & Richards, 1971). Each amide N atom was classified by its solvent-accessible area and exchange status. Ranges of 0.5–1 Å from 0 to 11 Å were used to bin the data. For each ASA bin, the fraction of residues with that ASA that had fully exchanged was calculated and a correlation coefficient was derived.

2.9. Global correlation fits

Essentially, we treated all 12 proteins as one very large protein or data set. Fully exchanged and total residues in binned ranges of the different HDX analysis parameters (depths, B factors, solvent accessibilities and hydrogen-bonding distances) were counted. The fully exchanged totals were then divided by the total number of residues (exchanged and non-exchanged) in that given bin and a percentage of fully exchanged residues in a given bin across all 12 proteins could then be calculated.

2.10. Programs used

Graphs, tables, plots and coefficients of determination (R^2 values from linear regression analysis of plotted data) were produced in *Microsoft Excel*. Figs. 8 and 9 and Supplementary Fig. 2 were generated using *PyMOL* (DeLano, 2002).

3. Results

Currently, there are 28 total proteins for which a neutron data set has been collected and reported in a peer-reviewed journal. As of March 2008, there are 22 available neutron coordinate sets for 14 different proteins deposited in the PDB. However, since only ten of these had refined occupancy values of the backbone amide D atoms, these are the structures that were used in the present analysis (see §2). Characteristics and features of the ten neutron crystal structures analyzed and discussed in this report are shown in Table 1. It is immediately clear that there is significant variance in some of the parameters for these structures, most importantly in resolution and overall data completeness. Given the small number of usable data sets in this type of analysis, it would be difficult to restrict ourselves by excluding any of the structures or by setting a resolution or completeness limit. However, it should be reiterated that the HDX analysis here was not performed on any structure refined at lower than 2.4 Å resolution. This variance, especially in data resolution, should be considered when interpreting the results and deriving conclusions, especially in terms of the deuterium occupancy and B-factor refinement. The other 18 proteins listed in Table 1 either do not have backbone deuterium-occupancy information listed in their coordinate files or their coordinate files have not been deposited in the PDB. All the crystals listed in Table 1 were soaked or vapor-exchanged with >90% D_2O for 21 d or more; however, those for which HDX information could be extracted had been incubated with D_2O for at least 30 d or longer (some for months to years).

3.1. Atomic depth, isotropic B factor, hydrogen-bonding distance and solvent accessibility

The interdependence between HDX and amide atomic depth, B factor, backbone amide hydrogen bonding and

solvent accessibility is presented in Table 2 and Figs. 2, 3, 4 and 5. Coefficients of determination (R^2 values) for each protein calculated from linear regression analysis of HDX plotted against these various parameters that are >0.7 are considered to be strong (values greater than 0.65 were rounded to 0.7 and were categorized as strong correlatives; see §2.1).

Atomic depth correlates strongly ($R^2 > 0.7$) with protection from H/D exchange for nine of the neutron structures. Atom depth is calculated using the *DPX* algorithm and is the distance (in angstroms) of the amide N atom from the nearest solvent-accessible atom (Pintar & Pongor, 2003). While it provides little information about residues on the protein surface (where depth = 0), it is a robust tool to describe the protein interior, allowing the folded globular protein to be treated and analyzed as discrete layers of atoms (Pintar *et al.*, 2003). For most neutron structures, there is a clear relationship of enhanced HDX protection for atoms that constitute interior layers of the proteins (those with DPX values of >1 Å). Of the 12 molecules examined, nine have R^2 values between 0.77 and 0.97, while three molecules, DHFR monomer B (DHFR-*B*; $R^2 = 0.1$), DsrD monomer B (DsrD-*B*; $R^2 = 0.35$) and rubredoxin ($R^2 = 0.04$), do not correlate (Fig. 2 and Table 2).

B factor seems to be related to HDX to a slightly lesser extent than depth. All proteins show positive correlation, as indicated by all of the slope trends being positive, supporting the notion that there is at least partial interdependence between an atom's magnitude of displacement (its vibrational motion) and its propensity to exchange (Fig. 3 and Table 2). In fact, five proteins (endothiapepsin, lysozyme, myoglobin and both RNase A structures) possess statistically strong correlation ($R^2 > 0.65$) between these two variables, showing that for these proteins amides that refine with high *B* factors tend to undergo HDX. XI possesses the weakest correlation between HDX and *B* factor ($R^2 = 0.12$).

Owing to the variability in the resolution of the structures and the number of different programs and routines that were used to refine isotropic temperature factors for the neutron structures analyzed here, an attempt was made to normalize the backbone amide *B* factors in order to increase the reliability of a cross-comparison of HDX correlation. We calculated the average backbone amide *B* factor and then calculated the ratio *B* factor/⟨*B* factor⟩ for each amide. This provides evidence of an amide's relative mobility, since ratios less than one suggests rigidity and ratios much greater than one suggest flexibility. Six of the 12 molecules possessed strong correlation for *B* factor/⟨*B* factor⟩ with respect to HDX (Table 2), with a higher percentage of exchanged residues having a higher *B*/⟨*B*⟩ ratio (showing a positive slope trend).

The R^2 values for HDX *versus* distance of hydrogen bonds to the amide vary widely from 0.92 to 0.0 (Fig. 4 and Table 2). Essentially, this is a measure of the interaction strength of the hydrogen bond: the shorter the distance, the stronger the bond and the greater the energy that would be necessary in order to break it (*i.e.* for an H/D exchange event to occur). We did not consider the hydrogen-bonding angle here; indeed, a surprising finding from a recent assessment of neutron structures by Niimura's group is that hydrogen-bonding geometries can vary significantly, with the donor/acceptor atom being non-collinear in the majority of hydrogen bonds observed (Niimura & Bau, 2008). As for hydrogen-bonding distance to the amide and its relationship to HDX, two proteins show strong correlation, lysozyme ($R^2 = 0.92$) and endothiapepsin (0.88), while another, BPTI (0.63), possesses moderate correlation. One of the important and striking features to note here is the slope trend of the plots. For half of the proteins this is positive, meaning that as hydrogen-bonding distances to the amide become longer a higher percentage of these amides have undergone HDX. However, for four other proteins the slope trend is negative. For the remaining two proteins, DHFR monomer *A* and myoglobin, the R^2 values are essentially zero, indicating no correlation at all (positive or negative) between HDX and hydrogen-bonding strength.

The weakest agreement is the correlation between the absolute solvent-accessible surface area (ASA) of the amide N atom and its tendency to undergo HDX (Fig. 5 and Table 2). By analysis of the R^2 values, only one protein (RNase A; 5rsa) shows a strong correlation ($R^2 = 0.68$) and three proteins [RNase A (6rsa), rubredoxin and DsrD-*A*] show moderate correlation of HDX to ASA. Three of the molecules (both DHFR monomers and DsrD-*B*) actually show no correlation between ASA and HDX. The HDX correlation data for all of the factors and all the neutron structures discussed above are summarized in Table 2.

Table 2
Structural parameters and correlation to backbone HDX.

Structure	PDB code	MW (kDa)	Depth R^2†	Depth Slope trend	B factor R^2	B factor Slope trend	$B/\langle B\rangle$ R^2	$B/\langle B\rangle$ Slope trend	Hydrogen-bonding distance R^2	Hydrogen-bonding distance Slope trend	Solvent accessibility R^2	Solvent accessibility Slope trend
DHFR-*A*	2inq	18	0.84	−	0.47	+	0.001	N/A	0.0003	N/A	0.37	−
DHFR-*B*	2inq	18	0.1	−	0.29	+	0.13	−	0.14	+	0.056	−
DsrD-*A*	1wq2	8.8	0.78	−	0.41	+	0.73	+	0.77	−	0.61	+
DsrD-*B*	1wq2	8.8	0.38	−	0.57	+	0.33	+	0.62	−	0.0003	−
Endothiapepsin	1gkt	33.8	0.87	−	0.88	+	0.99	+	0.88	+	0.08	+
Lysozyme	1lzn	14.3	0.96	−	0.8	+	0.78	+	0.92	+	0.28	+
Myoglobin	1l2k	17.2	0.66	−	0.78	+	0.71	+	0.01	N/A	0.23	+
Rubredoxin	1vcx	5.9	0.04	−	0.43	+	0.41	+	0.47	+	0.56	+
Xylose isomerase	2gve	43.3	0.86	−	0.31	+	0.05	+	0.12	−	0.37	+
Ribonuclease A	5rsa	13.7	0.97	−	0.77	+	0.67	+	0.42	−	0.68	+
Ribonuclease A	6rsa	13.7	0.91	−	0.83	+	0.85	+	0.08	+	0.63	+
BPTI	5pti	6.5	0.65	−	0.52	+	0.56	+	0.63	+	0.54	+
R^2 average			0.67		0.59		0.52		0.42		0.32	
Global R^2			0.92		0.29		0.13		0.92		0.19	

† R^2 is the correlation coefficient from linear regression analysis of HDX against the five parameters amide depth, amide *B* factor, normalized *B*-factor ratio, hydrogen-bonding distance and the amide solvent accessibility.

3.2. Secondary structure

For all the structures examined, the average deuterium occupancy for amides within β-strands is 0.42; for α-helices it is 0.57 and for unstructured regions it is 0.65. If both α-helices and β-strands are combined, the average occupancy for amides within secondary structure is 0.5; in every case apart from DsrD-*A* and DsrD-*B* the average occupancy for amides within secondary structures (0.45) is markedly lower than for amides within randomly structured regions (0.59) (see Supplementary Fig. 1[1]).

The notion of a microscopic hydrogen-protection factor (PF) has been elucidated by others (Bai *et al.*, 1993). Protection of backbone amides is often invoked as correlating with the type of secondary-structural element that the particular backbone amide is within, with α-helices and β-sheets demonstrated to be more protective elements than loops, turns and randomly structured regions. Enhanced van der Waals contacts (tighter packing), intra-chain hydrogen bonding (helices) and inter-chain hydrogen bonding (strands) are major contributing factors to observed HDX protection (Vendruscolo *et al.*, 2003).

A protection factor for a particular residue *i* within a folded protein can be calculated as $P_i^{exp} = k_i^{int}/k_i^{ex}$, comparing the intrinsic HDX rate (k_i^{int}) of residue *i* observed in an unstructured peptide dissolved in D_2O with the observed HDX rate (k_i^{ex}) of residue *i* in the folded protein (Hvidt & Nielsen, 1966; Bai *et al.*, 1993; Fersht, 1998; Best & Vendruscolo, 2006). Analogous to the kinetic and solution HDX information for proteins for which a neutron structure is available, we propose using the structurally derived 'relative protection factor' (RPF) with information from deuterium-occupancy refinement at backbone amide positions serving as the experimental HDX data. The criteria used for the classification of an exchanged *versus* an unexchanged amide are provided in §2.2. We have applied this to the ten neutron protein structures that have HDX information available. This exchanged fraction is denoted ss_{exch}/ss_{all}; in the case of those residues in unstructured regions of the protein (such as in randomly structured regions) this ratio is written $unstruc_{exch}/unstruc_{all}$. We define RPF_{ss} for a given type of secondary structure in a given neutron structure as the ratio of $unstruc_{exch}/unstruc_{all}$ to ss_{exch}/ss_{all},

$$\frac{unstruc_{exch}/unstruc_{all}}{ss_{exch}/ss_{all}}. \tag{1}$$

The assumption here is that the exchange propensity of an amide in a region devoid of secondary structure in a folded protein approximates the exchange propensity of an amide in an unfolded protein or in an unstructured peptide. Thus, the higher the RPF value, the greater the observed protection from HDX. Unstructured regions then have a defined and invariable RPF of 1.0. The RPF_{ss} values of ten proteins for which a neutron structure is available are presented in Table 3.

Table 3
Protection factor for backbone HDX correlated to secondary-structure type.

The relative protection factor (RPF) based on fully exchanged residues in a given secondary-structure type was calculated using the equation ($unstruc_{exch}/unstruc_{all}$)/($ss_{exch}/ss_{all}$), where $unstruc_{exch}$ and ss_{exch} are the number of residues within unstructured regions and secondary structures, respectively, whose amides have exchanged. $unstruc_{all}$ and ss_{all} are the total number of residues in unstructured regions and secondary structures, respectively. The global correlation of RPF to HDX was 0.92 for both secondary-structure types, where the global correlation (R^2 value) was calculated by linear regression analysis of the relationship between RPF for secondary structures and the refined deuterium occupancies of backbone amides in these secondary structures. The latter were measured by solving the equation $Docc_{uns}/(1/n)\sum_{i=1}^{n} Docc_i$.

Structure (PDB code)	Helix (H)	Strand (S)	H − S
DHFR-*A* (2inq)	1.159	1.405	−0.246
DHFR-*B* (2inq)	0.993	1.252	−0.259
DsrD-*A* (1wq2)	0.929	1.75	−0.821
DsrD-*B* (1wq2)	0.794	0.813	−0.019
Endothiapepsin (1gkt)	0.833	1.504	−0.671
Lysozyme (1lzn)	1.652	1.652	0
Myoglobin (1l2k)	1.174	None	NA
Rubredoxin (1vcx)	None	1.209	NA
Xylose isomerase (2gve)	1.265	1.586	−0.321
RNase A (5rsa)	1.259	1.574	−0.315
RNase A (6rsa)	1.221	1.324	−0.103
BPTI (5pti)	0.815	2.444	−1.629
Average	1.10	1.50	−0.4384
Standard deviation	0.26	0.41	0.4940547

Again, two of these (DHFR and DsrD) are crystallographic dimers, so a total of 12 coordinate sets are available for analysis. Clearly, β-sheet structure provides statistically significant protection from HDX: ten of the 12 molecules have a protection factor greater than one (in fact, for BPTI the RPF is 2.4!). On the other hand, for α-helices only six of the 12 molecules have a protection factor that is greater than one. In every case where a protection-factor comparison between secondary structures is possible (ten of the 12 molecules),
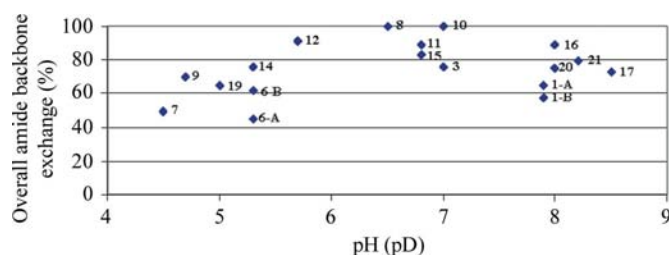

Figure 6
Amide backbone exchange plotted against crystal D_2O-soaking pH. The magnitude of backbone HDX in crystals is most adversely affected at acidic and even slightly basic pH (pD) levels. Proteins are labeled with their corresponding number in Table 1 (1-A is DHFR monomer *A*, 1-B is DHFR monomer *B*, 6-A is DsrD monomer *A* and 6-B is DsrD monomer *B*). Only 16 of the 25 proteins listed in Table 1 were analyzed. Two of the proteins (DHFR and DsrD) crystallize with two molecules in the asymmetric unit; thus, they are included here as four separate data points. Exchange data was not provided in the literature or could not be found in deposited structures for the nine proteins missing from this analysis. Limitations arising from resolution (lower than 2.5 Å) could very well prohibit this type of analysis.

amides within β-sheets are much more protected from exchange than those within helices. On a global scale (where exchange in all the secondary-structure types in all the proteins are considered), relative protection by secondary structure (RPF$_{ss}$) correlates well ($R^2 = 0.92$) with the exchange patterns of the neutron structures (RPF$_{Docc}$), as judged by the refined D occupancies of the backbone amide atoms from unstructured and structured (α-helix, β-strand or loop/turn) regions of the proteins. RPF$_{Docc}$ is measured using

$$\frac{\text{Docc}_{uns}}{\frac{1}{n}\sum_{i=1}^{N}\text{Docc}_i}. \qquad (2)$$

Visual evidence for the interdependence between HDX, secondary structure, atomic depth and B factor is provided in Supplementary Fig. 2. The ten neutron structures (12 total molecules) used in the HDX analysis have been drawn as C$^{\alpha}$ tubes that are color-coded based on backbone amide exchange (green indicates exchange has occurred, while red indicates the amide has resisted exchange). The tube width corresponds to the refined B factor of the backbone amide.

Another external factor influencing overall HDX in a crystal is the pH [or pD, which is +0.4 units greater than the pH (Glasoe & Long, 1960), presumably owing to the stronger covalent bonds that deuterium can form] of the D$_2$O-soaking solution. There are 19 crystal structures for which overall HDX data is known; the pH range is 4.5–8.5. Fig. 6 shows the trend for HDX *versus* pH of crystallization.

4. Discussion

4.1. HDX analysis of neutron protein structures: general findings

Backbone amide-exchange patterns will be influenced by multifarious factors. However, using an analysis of an existing subset of neutron protein structures, it is clear that the type of secondary structure and the depth within the protein are the major factors that can serve as general determinants of whether or not exchange will occur at a given backbone amide position. It must be restated that there are important limitations to both performing HDX analysis and to interpreting the results. These include the resolution of the data and the resolution limit imposed in refinement, both the overall completeness and the completeness in the highest resolution shell and the normally poor data-to-parameter ratio. From a statistical standpoint, another caveat is the available number of neutron structures at present that can be analyzed. The number of reported neutron structures is still frustratingly small; as of March 2008, only 22 have been deposited in the PDB. Of these, ten have detailed HDX information: this sample size represents perhaps the minimum cohort that can be used in a statistical analysis such as this one.

There are, of course, practical limitations on HDX occurring within protein crystals. These are environmental factors that the researcher can modulate, thus possibly influencing HDX in the crystal. One is the total incubation time in D$_2$O-based buffer. All the crystals listed in Table 1 for which HDX information is available (17 of the 28 listed) have been soaked for >30 d. The average extent of HDX for the five crystals that was soaked for the minimum time period tested (30) is 62%. So, for most proteins, this is likely to be a sufficient duration to allow significant backbone exchange to occur. Although temperature should play a role in the rate of exchange, it is most likely not a major factor in overall exchange when soaking experiments last for days to weeks. The most common temperature at which D$_2$O soaks are performed is ambient. However, the DHFR crystal used for ND experiments was soaked at for 30 d at 277 K (Bennett *et al.*, 2006); even at this temperature HDX has occurred for 65% and 57% of the amides in the two monomers in the asymmetric unit.

We have also considered the effect of pH of crystal growth on HDX (Fig. 6). Qualitatively, our analysis shows a moderate correlation of HDX with pH. However, the correlation between HDX and pH within our data set of 19 proteins is not as strong as that reported in solution-based studies. In particular, DHFR-*A* and DHFR-*B*, rubredoxin and XI can be considered to be outliers (Fig. 6). Interestingly, HDX seems to occur maximally within protein crystals between a pH of 6 and 7.5. At more basic pH (>8) there is a decrease in exchange. However, the average overall exchange of crystals soaked at or greater than pH 8 is 75%, compared with a 63% average for crystals soaked at or less than pH 5.5. At 49% HDX, endothiapepsin is the least exchanged; this is perhaps best explained because its soaking pH was 4.5 (Coates *et al.*, 2001), even though it was soaked in D$_2$O for 70 d. Another reason for this may be that it is a β-rich structure. Finally, HDX propensity seems to be independent of the solvent content of the crystal (Supplementary Fig. 3).

4.2. HDX and atomic depth

A previous study showed that amide depth correlates significantly better than solvent accessibility with HDX data measured from solution experiments (Chakravarty & Varadarajan, 1999). In fact, the average coefficient of determination (R^2 value) for depth reported in that analysis (0.61; Table 2 in Chakravarty & Varadarajan, 1999) is very similar to that calculated in the present report (0.62; Table 2). Depth calculations are a more robust descriptor of protein interiors than ASA values, which provide information about the protein surface and regions of the protein the solvent probe can actually access (Pintar *et al.*, 2003). Atomic depth values may ultimately serve as a sufficient prediction tool to determine the propensity of sites to exchange in the absence of neutron structural information. Furthermore, because the method of depth calculations discussed here is linked so intimately to distance to solvent-accessible atoms, it provides evidence that, at least in protein crystals, solvent penetration may be an important factor in a general mechanism for exchange. Of the 12 molecules considered in this analysis, nine possessed statistically strong correlation ($R^2 > 0.65$) between HDX and amide depth of burial.

A 'global' analysis of the correlation between HDX and amide depth is shown in Fig. 7. We treated the 12 different proteins as one data set, accumulating all exchanged residues from all the proteins into specified depth values. By summing all of the structures together, we also hope to dampen any effects from extreme variance in the data (outliers), some of which could be attributed to limitations of the data owing to
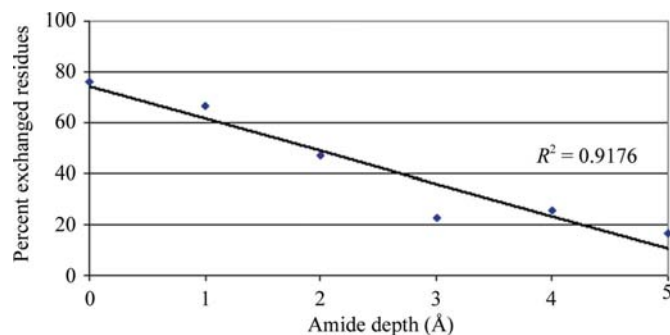

Figure 7
A global correlation plot of HDX against amide atomic depth. A cumulative total percentage of fully exchanged residues at particular depths from all eight proteins in this analysis were calculated and then plotted. The R^2 value shown was calculated by linear regression fitting. The atom depth of the amide is the parameter with the highest correlation to HDX as determined from this global analysis; indeed, it is the only one that shows statistically strong correlation. Plots of global correlation for the other parameters considered in this analysis can be found in Supplementary Fig. 4. All have R^2 values of <0.3.
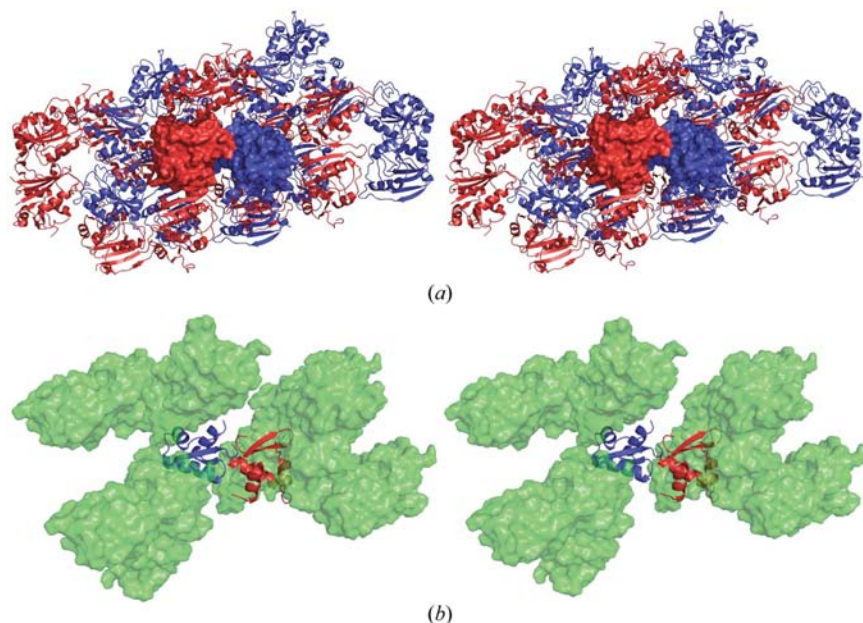

Figure 8
Unequal crystal-packing environments for the two molecules in the (a) DHFR and (b) DsrD asymmetric units. (a) Stereoview of the crystal packing in the DHFR asymmetric unit. The DHFR origin molecules are shown as surface representations (monomer A, red; monomer B, blue). Only symmetry-related (6_1 on c) molecules that are within van der Waals contact range (<5 Å) of any atom in either of the origin molecules are shown (cartoon representations; colored as the origin molecules). (b) Stereoview of the crystal packing in the DsrD asymmetric unit. The DsrD origin molecules are shown as cartoons colored as the origin molecules in (a). Again, only symmetry-related (2_1 on abc) molecules within van der Waals contact range of the origin molecules are shown (slightly transparent green surface representations).

resolution and/or completeness. The global R^2 for atomic depth is 0.92 (Fig. 7). We performed the same calculation for the other three factors (amide B factor, hydrogen-bonding distances to the amide and amide ASA); this is shown in Supplementary Fig. 4. For the amide B factor the global R^2 is 0.29, for hydrogen-bonding strength it is 0.92 and for ASA it is 0.19. A caution should be added when considering the correlation with hydrogen-bond strength. Supplementary Fig. 5 shows that even at the shortest hydrogen-bonding distance of 2.4 Å, 56% of the amides undergo HDX. Hence, short or strong hydrogen bonds do not completely inhibit HDX, but rather modulate it. Furthermore, for hydrogen-bonding strength, only two of the proteins (endothiapepsin and lysozyme) show this strong interdependence (with the expected positive slope trend). In contrast, atomic depth not only correlates globally with HDX; on an individual protein basis it correlates strongly with HDX for ten of the 12 proteins analyzed.

Two of the protein molecules for which HDX behavior does not correlate to atomic depth are monomer B of DHFR ($R^2 = 0.13$) and monomer B of DsrD ($R^2 = 0.38$) (Fig. 2). (The other protein that shows poor correlation to depth is rubredoxin, which is discussed in a later section.) This is surprising because the HDX patterns of the other monomer in the asymmetric unit of both of these crystals correlate quite strongly with depth ($R^2 = 0.72$ for DHFR-A; $R^2 = 0.77$ for DsrD-A). Interestingly, overall HDX itself (and even the pattern of exchange) differs somewhat between the two monomers in the asymmetric unit for both crystals. The C^α r.m.s.d. between the two monomers is 0.5 Å (DHFR) and 0.7 Å (DsrD); however, local regions of conformational differences exist between the monomers in both cases. For DHFR, these differences mainly occur at important regulatory-loop regions, namely the Met20 and $\beta(F-G)$ loops as previously mentioned (Bennett et al., 2006). In monomer A, the Met20 loop folds in towards the cofactor-binding site, partially occluding it, and away from the $\beta(F-G)$ loop. This results in extensive exchange, significantly more than in monomer B. This is because in monomer B the Met20 and $\beta(F-G)$ loops each adopt different conformations; the Met20 loop moves to a closed position. This results in the loops packing against one another and forming multiple hydrogen bonds. The amide-exchange heterogeneity between the two molecules of the asymmetric unit, exemplified by the differing HDX patterns in these structurally distinct loop regions, may allow one to consider these monomers as being so different that it is not surprising that the depth correlation is not equal for both. For DHFR-B, the HDX correlation to depth

essentially breaks down at depths greater than 4 Å, with half of the amides at these depths being exchanged. It should be stressed that the sample size at these depths is relatively small; eight amides have depths of >4 Å out of a total of 149 amides in DHFR-*B*.

The explanation for DsrD monomer *B* is more complicated. Here, correlation is poor because as depth increases up to 3 Å the extent of HDX is actually greater. There is only one amide with a depth beyond 3 Å in either monomer (again, a very small sample size) and it is not exchanged in both. The region of largest structural difference between the two monomers constitutes residues 14–22. In the X-ray structure this region is a short helical segment in monomer *A*, while in monomer *B* it adopts a randomly ordered structure. However, in the neutron structure, residues within this region are not modeled for both monomers (15–19 in *A*; 16–17 in *B*). Of the residues for which there is amide-exchange information (seven between the two monomers), all but one (Ser14 in *A*) have resisted HDX and all but one (Phe22 in *B*) have a depth of 0, indicating that they are at the surface.

What makes DHFR and DsrD stand out from our data set of ten proteins is the fact that these are the only two proteins to crystallize with two monomers per crystallographic asymmetric unit; all the other crystals have one monomer per crystallographic asymmetric unit. When only one monomer is present in the asymmetric unit, the crystal packing around any molecule within the crystallographic unit cell is identical owing to the propagation of crystallographic symmetry. However, this is not the case for the two molecules within the asymmetric unit. For example, monomer *A* in both DHFR and DsrD will have a different crystal-packing configuration compared with monomer *B* (Fig. 8). In this case, the crystal packing can influence the HDX behavior of the two molecules differently. This might be one of the reasons for the disparity in the HDX patterns and in the correlation of HDX to depth that exists between the two molecules of DHFR and DsrD.

On the other hand, the RNase A structures 5rsa and 6rsa that crystallized in the same unit cell with one molecule in the asymmetric unit, hence having identical crystal contacts, correlate well with each other's HDX patterns ($R = 0.92$), providing a useful benchmark on the reproducibility of the data that were used for the analysis.

4.3. Depth: nearest solvent-accessible atom *versus* nearest water molecule approaches

Interestingly, a previous HDX study based on the ND structure of trypsin revealed that greater than 90% of non-exchanged amides were located at least 4 Å from the protein surface (Kossiakoff, 1982). Of the fully exchanged amides, ~23% were also located 4 Å or more from the protein surface. However, in this study an alternate approach was used to derive depth. Here, the authors defined depth as the distance between the amide N atom and the nearest bulk-solvent interface; this value could be directly determined using the nuclear density maps. Unfortunately, we were unable to derive the atomic depth of trypsin using the *DPX* algorithm as the

deposited coordinates do not have refined occupancies binned in the ranges listed in the primary citation (Kossiakoff, 1982). However, for nine of the 12 molecules included in the present analysis we found strong correlation ($R^2 > 0.65$) between the depth of an amide N atom as calculated by *DPX* and its propensity to undergo HDX. Additionally, on a global scale, where all the residues of all 12 molecules are binned into given ranges, atomic depth is the only parameter of the four we tested that correlates strongly ($R^2 = 0.92$) both on a global scale (Fig. 7) and on an individual protein basis (Fig. 2). This value is significantly higher than the average R^2 calculated from all the depth correlation plots of the individual proteins (0.65; Table 2). The fit is markedly better for the global correlation because the data set is much larger and skewing effects from outliers (such as DHFR-*B*, DsrD-*B* and rubredoxin) are dampened in this type of analysis. Further support for the correlation of HDX *versus* depth comes from an independent study on lysozyme, where both ND and NMR data measured from triclinic crystals were compared (the crystals were dissolved in a D_2O-based buffer prior to the beginning of the NMR experiment; Pedersen *et al.*, 1991). The authors employed a 'nearest hypothetical water molecule' approach to define atom depth (Pedersen *et al.*, 1991). Their study revealed that in addition to secondary structure, an important factor in explaining exchange behavior is the depth of the amide N atom, as measured by its distance from the nearest hypothetical water molecule (Pedersen *et al.*, 1991). Indeed, in our study the R^2 value determined for lysozyme relating depth to HDX approaches 1 ($R^2 = 0.96$), suggesting particularly strong correlation between the two variables. It should also be noted that good agreement was observed in the assignment of exchanged amides between the two methods (ND and NMR): 113 of the 121 amides for which HDX could be assigned agreed. Furthermore, a study of HDX in sperm whale myoglobin as measured by NMR (Cavagnero *et al.*, 2000) reveals very good agreement in identifying the protected amides in the NMR analysis and in the ND structure (Ostermann *et al.*, 2002). Of the seven most highly protected amides (RPF > 7, calculated from the ratio k^{exch}/k^{int}, similar to described above for secondary structure) identified in the NMR study (Cavagnero *et al.*, 2000), all but one refine with a deuterium occupancy of 0 in the ND structure, indicating strong protection from exchange in the myoglobin crystals (Ostermann *et al.*, 2002).

4.4. Depth, HDX and molecular weight

While the size of the protein does not directly correlate to HDX propensity (Supplementary Fig. 6), it does seem to influence how well the atomic depth values correlate with HDX. Indeed, a limitation of using this type of depth analysis in correlation to HDX propensity is demonstrated by rubredoxin (Kurihara *et al.*, 2004), which shows the poorest correlation ($R^2 = 4.2\%$) of all the protein structures tested. It is the smallest protein included in the analysis (53 residues; 5.9 kDa); nearly 90% of backbone amides have exchanged and only one residue has a depth beyond 3 Å. In other words,

in terms of the depth calculations performed here, the vast majority of the polypeptide is at or very near the surface. Admittedly, when the *DPX* algorithm is used for depth calculations some information is lost for surface residues because any solvent-accessible atom is given a depth of 0 (Pintar & Pongor, 2003; Pintar *et al.*, 2003). Even considering this, the suggestion is that rubredoxin, while possessing secondary-structural elements, may be devoid of a true protein core.

In contrast to the very small proteins included in this analysis (such as rubredoxin), the two largest proteins, XI (43.3 kDa) and endothiapepsin (33.8 kDa), have HDX patterns that correlate strongly with the atom-depth calculations (with linear R^2 values of 0.85 and 0.87, respectively). Each of these proteins has an interior structural unit that is significantly rich in β-strands (antiparallel β-sandwiches in endothiapepsin and a partial β-barrel in XI); most of the residues that comprise these strands are resistant to HDX (refined D occupancy values <0.2) and may serve as core domains.

4.5. HDX and B factors

The correlation between *B* factors and HDX is not conclusive. Our results show that there is no strong global correlation between these two parameters ($R^2 = 0.29$). However, there are shortcomings in conducting a global correlation analysis using *B* factors derived from structures that have been refined using different programs. For instance, it is known that *CNS* (Brünger *et al.*, 1998) tends to give larger *B* factors compared with other programs such as *REFMAC* (Murshudov *et al.*, 1997) and *SHELX* (Sheldrick, 2008). To avoid this problem, we derived an averaged correlation value, where the coefficient of determination between HDX and *B* factor for each structure was averaged. Based on this method, the average correlation for HDX against *B* factor was 0.57 (Table 2). Seven of the 12 molecules analyzed possessed R^2 values that were greater than 0.5, thus showing moderate correlation. Additionally, we determined the variance of amide *B* factors from an average *B* factor, using the ratio $B/\langle B \rangle$, and then analyzed the correlation of this ratio with HDX. This gave very similar results to the *B*-factor analysis, with six of the 12 molecules showing strong correlation (positive slope trends).

It is not entirely surprising that there is not a correlation between HDX and *B* factor. For instance, Karplus and coworkers have reported that the *B* factors for α-lactalbumin, as derived from its X-ray crystal structure, do not correlate strongly with experimentally determined amide-exchange protection factors, especially in its C-terminal domain (Vendruscolo *et al.*, 2003). Indeed, they could identify a vast ensemble of conformational states that the protein could sample only rarely but that were sufficient to allow HDX to occur. A number of these fluctuating regions possessed relatively low *B* factors. Based on these findings and our present analysis, amide *B* factors correlate only moderately to HDX propensity.

4.6. HDX and hydrogen-bonding strength

We examined the relationship between the lengths of hydrogen bonds made to backbone amides and their propensity to undergo HDX. Previously, Kossiakoff and coworkers had found a correlation between HDX patterns and hydrogen bonding for trypsin (Kossiakoff, 1982) and subtilisin (Kossiakoff *et al.*, 1991). However, it was emphasized that this correlation was most likely to be a consequence of the observation that regular hydrogen bonding within secondary structures such as helices and strands seemed to suppress HDX (Kossiakoff, 1982; Kossiakoff *et al.*, 1991). Our study, which was conducted with several more neutron protein structures, shows that on an individual protein basis HDX does not correlate with the strength of backbone hydrogen bonding. In fact, for four of the protein molecules tested the slope of the linear regression trend line is negative (Fig. 4 and Table 2). For these molecules, amides forming stronger intramolecular hydrogen bonds (bonds formed with other protein atoms only) tend to be fully exchanged. This is an interesting observation as it would be expected that these amides would resist exchange, especially if they reside within secondary structure that is stabilized by hydrogen bonding. If amides form hydrogen bonds to nonsolvent atoms yet still undergo exchange, one then assumes that these bonds can be broken to allow transient interaction with D_2O so as to catalyze an exchange event. This may lend evidence that local and subglobal unfolding of secondary structure occurs even in the crystalline state. This is a special case of the aforementioned phenomenon termed protein breathing (Englander *et al.*, 1972, 1980), a mechanism by which the polypeptide chain can partially unfold, sampling new conformations and folding space. A consequence of these structural rearrangements is that backbone amides within this region come into temporary yet energetically favorable contact with D_2O and can then undergo exchange (Kossiakoff, 1982).

4.7. HDX and solvent accessibility

We found no interdependence between backbone exchange and the ASA of the amide. One can perhaps gain little information about binding sites and important clefts and cavities in a protein (even near the surface) with solvent-accessibility values simply because the probe itself used in accessibility calculations (normally a 1.4 Å probe radius to mimic a solvent molecule, as in Lee & Richards, 1971) may not be able to physically access the site owing to the steric effect imposed by the protein's structure (Pintar *et al.*, 2003).

4.8. HDX and secondary structure

For every molecule included in our analysis (except for both monomers of DsrD), backbone amides within α-helical and/or β-strand structure are significantly protected from exchange relative to unstructured regions within the proteins. We measured this by devising a relative protection factor (RPF) for a secondary-structure type in a given protein, where the fraction of exchanged residues in unstructured regions is divided by the fraction of exchanged residues in secondary-

structural elements. For all the proteins tested, we also determined the average deuterium occupancy (Docc) of amides within unstructured regions and within regions of secondary structure. The average Docc for secondary structures is 0.5, while for unstructured regions it is 0.65 (Supplementary Fig. 1). As has been observed with many proteins using other techniques, β-sheet structure is the most protective secondary structure against HDX (Raschke & Marqusee, 1998; Fersht, 1998; Vendruscolo *et al.*, 2003; Yamamoto *et al.*, 2004; Polshakov *et al.*, 2006). Furthermore, as β-strand content increases, overall HDX diminishes in a somewhat linear fashion.

The β-rich regions (and even helical elements within or near the core; Zhang & Peng, 2000) are highly protected from solvent intrusion and HDX (Creighton, 1993; Richardson & Richardson, 2002), mainly because distortions of β-strands involve torsion changes that do not normally cause breakage of interstrand hydrogen bonding. A similar degree of distortion in helical structure would normally cause temporary breaking of hydrogen bonds, making HDX possible (Wlodawer & Sjolin, 1982; Kossiakoff, 1982; Kossiakoff *et al.*, 1991). These regions may constitute minimal folding units (Lesk & Rose, 1981) which form early in the overall folding process for the given protein (Englander, 2000). A previous study on lysozyme, in which exchange in crystals (monitored by ND) and exchange in solution (monitored by 2D-NMR) were compared, determined that an important factor influencing exchange behavior is the secondary-structure environment of the backbone amide (Pedersen *et al.*, 1991). Specifically, it was argued that participation of the amide in secondary-structure interactions suppresses exchange. This is also what we have found to exist for nearly all the protein molecules analyzed in this report. It is important to note that both monomers of DsrD break this trend. We believe this is because unstructured regions of DsrD participate in crystal contacts to a much greater extent than residues in secondary structure. The effect of crystal packing may very well be the reason the amides within unstructured regions invariably have lower Docc values than amides within helices or strands.

4.9. HDX and the protein core

Core regions of proteins tend to resist exchange (Figs. 2, 7, 9 and Supplementary Fig. 2) and may serve as minimal folding domains (Fersht, 1998; Li & Woodward, 1999); these are seed structures that are formed near the beginning of the folding pathway and help to funnel the protein towards a stable globular structure (Englander *et al.*, 1997; Dill & Chan, 1997). Recently, a new microscopic folding mechanism known as the zip-and-assembly method (ZAM) has been suggested by Dill and coworkers (Ozkan *et al.*, 2007). Local folding of short regions of polypeptide occur first on fast time scales, providing a catalytic template for surrounding regions to fold and resembling a zipper; the locally assembled folded structural units then culminate into a globally folded molecule (Dill *et al.*, 2007). These zipper regions are essentially the stable nuclei noted above and are likely to be components of hydrophobic

clusters in the core of the native protein. Therefore, the zipper regions may be highly resistant to HDX. Chakravarty & Varadarajan (1999) defined the hydrophobic core of a protein as residues within 1 Å of the deepest residue. From their atomic depth analysis of four different proteins, they found that up to 60% of the core residues have amide-exchange protection factors of >100 (Chakravarty & Varadarajan, 1999). Using DPX, we could determine the deepest backbone amide in each of the 12 molecules investigated here. We then calculated the average from these deepest amide values, which turns out to be 4.8 Å. For the proteins evaluated in this analysis, 74% of backbone amides at distances of >3.8 Å (within 1 Å of the deepest average residue) from the nearest solvent-accessible surface are not exchanged. Perhaps this distance ($\sim$3.8 Å) from the surface serves as a general boundary for the protein hydrophobic core, with some expected variability of this value based on protein size (for example, rubredoxin, myoglobin and DsrD have no clusters that have a depth of $\sim$3.8 Å).

Using the depth of $\sim$3.8 Å, we have determined the hydrophobic core of each of the proteins as clusters of residues that are at or greater than this depth. Based on these criteria, we have assigned the core for six molecules as shown in Fig. 9. Both RNase A molecules, XI, lysozyme and DHFR have single-clustered cores, while endothiapepsin has a triple-clustered core. In the two DHFR monomers the clusters constituting the core are far removed; only three residues overlap. This difference is a consequence of the aforementioned conformational changes between the partially occluded and closed DHFR molecules. The structure of endothiapepsin can be characterized as being composed of two rigid bodies: one is constituted of residues −2 to 189 and 303 to 326, with the other consisting of residues 190–302 (Sali *et al.*, 1989, 1992). Interestingly, from our analysis of its HDX/depth patterns, endothiapepsin appears to possess three hydrophobic cores (Fig. 9): two of these clusters (left and middle in Fig. 9) reside within the first rigid-body domain listed above, while the third cluster (right in Fig. 9) resides in the second rigid body.

For the neutron structures analyzed here, there are 13 amides at depths greater than 3.8 Å which have undergone HDX. They have done so possibly owing to a combination of factors: (i) they are either not within secondary structure or are at the termini of secondary-structure elements and/or (ii) they are adjacent to a potential solvent channel within van der Waals contact or hydrogen-bonding distance of a D_2O. We can confirm both these factors are plausible based on our analysis of the neutron structures for endothiapepsin (Coates *et al.*, 2001), DHFR (Bennett *et al.*, 2006) and XI (Katz *et al.*, 2006). [There are no exchanged amides at depths greater than 3.8 Å for DsrD (Chatake *et al.*, 2003), lysozyme (Bon *et al.*, 1999), myoglobin (Ostermann *et al.*, 2002) or either of the RNase A molecules (Wlodawer & Sjolin, 1982; Wlodawer *et al.*, 1983; Borah *et al.*, 1985)]. For example, in XI Glu217 is exchanged yet is 4.8 Å from the nearest solvent-accessible atom. However, a D_2O molecule is 2.7 Å from the amide N atom (Katz *et al.*, 2006), possibly explaining how exchange could

occur at such an interior site. Also, it is at the terminus of a helix and has not formed any hydrogen bonds to protein atoms. Further examples are the six amides within the hydrophobic core of DHFR (at depths of >3.8 Å) but which have undergone HDX; these are Ser61 (monomer *A*), Ile41 (*B*), Met42 (*B*), Val93 (*B*), Leu104 (*B*) and Tyr111 (*B*). All of these amides are within or adjacent to residues which are extremely conserved among DHFRs. All residues except for Tyr111 (*B*) are at the termini of secondary-structure elements, regions which may be prone to localized unfolding, thus allowing temporary D_2O contact. Tyr111 resides in the middle of β-strand *F*; its amide has a very low *B* factor (8.6 Å^2) and a depth of 4.8 Å, yet its refined deuterium occupancy is 0.94. However, we observe a nuclear density difference peak (>3σ) located within hydrogen-bonding distance ($\sim$2.6 Å) of the Tyr111 amide; this could be a D_2O molecule. Additionally, most of these residues, Ser61 (*A*), Ile41 (*B*), Met42 (*B*) and Val93 (*B*), have been implicated, based on results from molecular-dynamics simulations, multidimensional NMR spectroscopy (Boehr *et al.*, 2006) and kinetic studies of mutants (Agarwal *et al.*, 2002), to be part of a collaborative network of residues responsible for 'coupled promoting motions' in the enzyme, a mechanism by which disparate parts of the molecule can communicate and impose long-range regulation of ligand binding and catalytic events at the active site. Therefore, it may be expected and even necessary for these residues to be highly dynamic, if indeed they are a part of an allosteric network and/or critical residues for protein breathing patterns. Even in the crystalline state, the secondary structures in which these residues reside may undergo regional melting. Similar to findings from HDX experiments for proteins conducted in the solution state, local unfolding (and possibly unfolding of whole domains) is then also an important factor for explaining exchange in protein crystals (Englander *et al.*, 1972, 1980).

5. Conclusions

From evaluation of 12 neutron protein crystal structures, we have identified an interdependence of secondary structure and atomic depth with HDX: there is significant reduction of exchange within secondary structure (especially β-strands, as
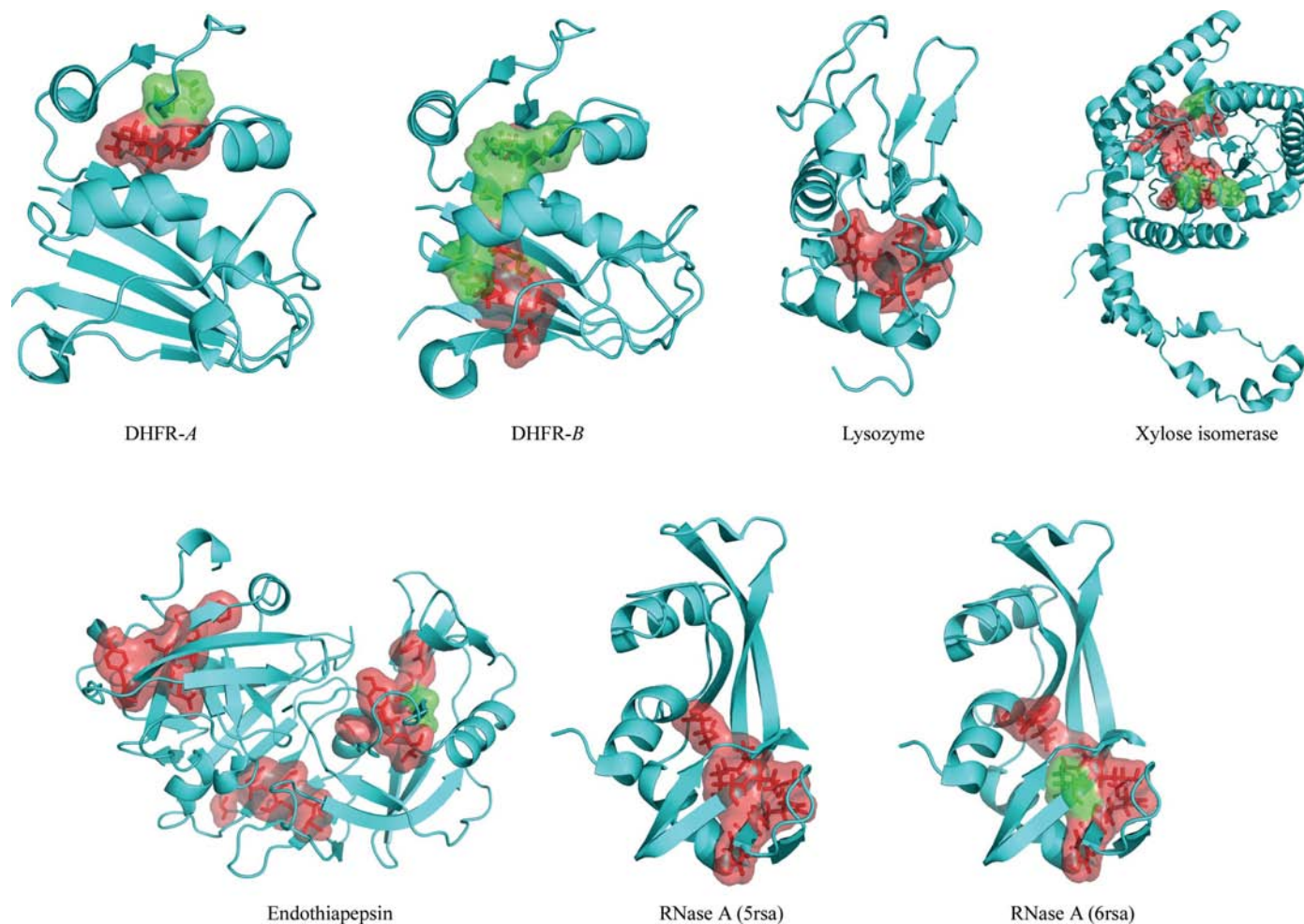

Figure 9
Hydrophobic cores identified from HDX data and depth calculations of the protein neutron structures analyzed here. Proteins are shown in cyan as cartoon representations. Residues which have backbone amides with depths of >3.8 Å are shown as semi-transparent surfaces and with side chains. A core is considered to be two or more residues at sufficient depth. Exchanged core residues are colored green, whereas non-exchanged core residues are shown in red. Most proteins possess a single contiguous core that is highly resistant to HDX. Endothiapepsin seems to have three well defined core units.

Bennett *et al.* · Amide exchange in neutron structures

observed by amide backbone deuterium occupancies of <0.2) and at protein cores (as defined by atomic depth calculations). On a whole-molecule level, what is governing exchange protection within protein crystals is both the incapability of local and subglobal unfolding events (Englander, 2000) to bring deeply buried amides out into solvent and the inability of solvent to gain access to these regions. Thus, proteins can be envisaged as molecules with distinct layers based on depth; essentially, HDX protection increases in a gradual linear fashion from the surface to the core. Exchange occurring within protein cores (those amides with depths near 3.8 Å or greater) happens owing to proximity to solvent channels and to observed D_2O molecules (Katz et al., 2006) and also owing to fluctuation of local and subglobal structure, as noted for several residues in DHFR that have been hypothesized to compose a dynamical allosteric network (Agarwal et al., 2002). It may be that secondary and/or tertiary structures within core layers are identical to initial localized folding units in the hierarchic (Lesk & Rose, 1981) and/or ZAM (Dill et al., 2007) folding pathways. As one proceeds into the core layers, protein architecture is dominated by relatively 'dry' secondary-structural units. Indeed, save for a few well ordered D_2O molecules observed within hydrophobic core regions in recent ND work on lysozyme (Bon et al., 1999) and DHFR (Bennett et al., 2006), these secondary structures tend to be dehydrated and compact. Structure within protein cores, often composed of small hydrophobic clusters (Ragona et al., 1997) such as those identified here by the depth method (Fig. 9), is most likely to be relatively rigid, so much so that it disallows even transient D_2O contact with most of the backbone amides. Additionally, the deeply buried amides will be protected because the unfolding of core units for proteins in the crystalline state is energetically unfavorable.

References

Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T., Benkovic, S. J. & Hammes-Schiffer, S. (2002). Proc. Natl Acad. Sci. USA, **99**, 2794–2799.
Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Proteins, **17**, 75–86.
Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Proteins, **20**, 4–14.
Bennett, B. C., Langan, P., Coates, L., Mustyakimov, M., Schoenborn, B., Howell, E. E. & Dealwis, C. G. (2006). Proc. Natl Acad. Sci. USA, **103**, 18493–18498.
Bennett, B. C., Meilleur, F., Myles, D. A. A., Howell, E. E. & Dealwis, C. G. (2005). Acta Cryst. D**61**, 574–579.
Berman, H. M. et al. (2002). Acta Cryst. D**58**, 899–907.
Best, R. B. & Vendruscolo, M. (2006). Structure, **14**, 97–106.
Blakeley, M. P., Kalb, A. J., Helliwell, J. R. & Myles, D. A. (2004). Proc. Natl Acad. Sci. USA, **101**, 16405–16410.
Blakeley, M. P., Ruiz, F., Cachau, R., Hazemann, I., Meilleur, F., Mitschler, A., Ginell, S., Afonine, P., Ventura, O. N., Cousido-Siah, A., Haertlein, M., Joachimiak, A., Myles, D. & Podjarny, A. (2008). Proc. Natl Acad. Sci. USA, **105**, 1844–1848.
Blum, M.-M., Koglin, A., Rüterjans, H., Schoenborn, B., Langan, P. & Chen, J. C.-H. (2007). Acta Cryst. F**63**, 42–45.
Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2006). Science, **313**, 1638–1642.
Bon, C., Lehmann, M. S. & Wilkinson, C. (1999). Acta Cryst. D**55**, 978–987.
Borah, B., Chen, C. W., Egan, W., Miller, M., Wlodawer, A. & Cohen, J. S. (1985). Biochemistry, **24**, 2058–2067.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Cryst. D**54**, 905–921.
Budayova-Spano, M., Bonneté, F., Ferté, N., El Hajji, M., Meilleur, F., Blakeley, M. P. & Castro, B. (2006). Acta Cryst. F**62**, 306–309.
Cavagnero, S., Theriault, Y., Narula, S. S., Dyson, H. J. & Wright, P. E. (2000). Protein Sci. **9**, 186–193.
Chakravarty, S. & Varadarajan, R. (1999). Structure Fold. Des. **7**, 723–732.
Chalmers, M. J., Busby, S. A., Pascal, B. D., He, Y., Hendrickson, C. L., Marshall, A. G. & Griffin, P. R. (2006). Anal. Chem. **78**, 1005–1014.
Chatake, T., Mizuno, N., Voordouw, G., Higuchi, Y., Arai, S., Tanaka, I. & Niimura, N. (2003). Acta Cryst. D**59**, 2306–2309.
Cheng, X. & Schoenborn, B. P. (1990). Acta Cryst. B**46**, 195–208.
Cipriani, F., Castagna, J. C., Wilkinson, C., Oleinek, P. & Lehmann, M. S. (1996). J. Neutron Res. **4**, 79–85.
Clarke, J. & Itzhaki, L. (1998). Curr. Opin. Struct. Biol. **9**, 112–118.
Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. & Cooper, J. B. (2001). Biochemistry, **40**, 13149–13157.
Collaborative Computational Project, Number 4 (1994). Acta Cryst. D**50**, 760–763.
Creighton, T. E. (1993). Proteins: Structures and Molecular Properties, 2nd ed. New York: W. H. Freeman & Co.
Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H. J., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). J. Chem. Soc. Faraday Trans. **93**, 4305–4312.
DeLano, W. L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.org.
Dill, K. A. & Chan, H. S. (1997). Nature Struct. Biol. **4**, 10–19.
Dill, K. A., Ozkan, S. B., Weikl, T. R., Chodera, J. D. & Voelz, V. A. (2007). Curr. Opin. Struct. Biol. **17**, 342–346.
Englander, S. W. (2000). Annu. Rev. Biophys. Biomol. Struct. **29**, 213–238.
Englander, S. W., Calhoun, D. B., Englander, J. J., Kallenbach, N. R., Liem, R. K., Malin, E. L., Mandel, C. & Rogero, J. R. (1980). Biophys. J. **32**, 577–589.
Englander, S. W., Downer, N. W. & Teitelbaum, H. (1972). Annu. Rev. Biochem. **41**, 903–924.
Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). Protein Sci. **6**, 1101–1109.
Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Curr. Opin. Struct. Biol. **6**, 18–23.
Fersht, A. R. (1998). Structure and Mechanism in Protein Science. New York: W. H. Freeman & Co.
Fisher, S. Z., Anderson, S., Henning, R., Moffat, K., Langan, P., Thiyagarajan, P. & Schultz, A. J. (2007). Acta Cryst. D**63**, 1178–1184.
Glasoe, P. & Long, F. (1960). J. Am. Chem. Soc. **64**, 188–190.
Habash, J., Raftery, J., Nuttall, R., Price, H. J., Wilkinson, C., Kalb (Gilboa), A. J. & Helliwell, J. R. (2000). Acta Cryst. D**56**, 541–550.
Habash, J., Raftery, J., Weisberger, S., Cassetta, A., Lehmann, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. W. & Helliwell, J. R. (1997). J. Chem. Soc. Faraday Trans. **93**, 4313–4317.

Hanson, B. L., Langan, P., Katz, A. K., Li, X., Harp, J. M., Glusker, J. P., Schoenborn, B. P. & Bunick, G. J. (2004). *Acta Cryst.* D**60**, 241–249.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Cryst.* D**61**, 1413–1417.

Hubbard, S. & Thornton, J. (1996). *NACCESS* v.2.2.1 – *Atomic Solvent Accessible Area Calculations.* http://www.bioinf.manchester.ac.uk/naccess/.

Hvidt, A. & Linderstrom-Lang, K. (1954). *Biochim. Biophys. Acta*, **14**, 574–575.

Hvidt, A. & Nielsen, S. O. (1966). *Adv. Protein Chem.* **21**, 287–386.

Hvidt, A. & Pedersen, E. J. (1974). *Eur. J. Biochem.* **48**, 333–338.

Ishikawa, T., Chatake, T., Ohnishi, Y., Tanaka, I., Kurihara, K., Kuroki, R. & Niimura, N. (2008). *Chem. Phys.* **345**, 152–158.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Katz, A. K., Li, X., Carrell, H. L., Hanson, B. L., Langan, P., Coates, L., Schoenborn, B. P., Glusker, J. P. & Bunick, G. J. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8342–8347.

Kossiakoff, A. A. (1982). *Nature (London)*, **296**, 713–721.

Kossiakoff, A. A. & Spencer, S. A. (1981). *Biochemistry*, **20**, 6462–6474.

Kossiakoff, A. A., Ultsch, M., White, S. & Eigenbrot, C. (1991). *Biochemistry*, **30**, 1211–1221.

Kovalevsky, A. Y., Chatake, T., Shibayama, N., Park, S.-Y., Ishikawa, T., Mustyakimov, M., Fisher, S. Z., Langan, P. & Morimoto, Y. (2008). *Acta Cryst.* F**64**, 270–273.

Kurihara, K., Tanaka, I., Chatake, T., Adams, M. W., Jenney, F. E. Jr, Moiseeva, N., Bau, R. & Niimura, N. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 11215–11220.

Langan, P., Greene, G. & Schoenborn, B. P. (2004). *J. Appl. Cryst.* **37**, 24–31.

Lawson, C. L. & Chin, A. S. (2003). *Trans. Am. Crystallogr. Assoc.* **38**.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Lesk, A. M. & Rose, G. D. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4304–4308.

Li, R. & Woodward, C. (1999). *Protein Sci.* **8**, 1571–1590.

Li, X., Langan, P., Bau, R., Tsyba, I., Jenney, F. E., Adams, M. W. W. & Schoenborn, B. P. (2004). *Acta Cryst.* D**60**, 200–202.

Loh, S. N., Prehoda, K. E., Wang, J. & Markley, J. L. (1993). *Biochemistry*, **32**, 11022–11028.

Maier, C. S. & Deinzer, M. L. (2005). *Methods Enzymol.* **402**, 312–360.

Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). *Biochemistry*, **11**, 150–158.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Myles, D. A. (2006). *Curr. Opin. Struct. Biol.* **16**, 630–637.

Niimura, N. & Bau, R. (2008). *Acta Cryst.* A**64**, 12–22.

Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J. C., Cipriani, F., Hoghoj, P., Lehmann, M. S. & Wilkinson, C. (1997). *Nature Struct. Biol.* **4**, 909–914.

Ostermann, A., Tanaka, I., Engler, N., Niimura, N. & Parak, F. G. (2002). *Biophys. Chem.* **95**, 183–193.

Ozkan, S. B., Wu, G. A., Chodera, J. D. & Dill, K. A. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 11987–11992.

Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M. & Redfield, C. (1991). *J. Mol. Biol.* **218**, 413–426.

Pintar, A., Carugo, O. & Pongor, S. (2003). *Trends Biochem. Sci.* **28**, 593–597.

Pintar, A. & Pongor, S. (2003). *Bioinformatics*, **19**, 313–314.

Polshakov, V. I., Birdsall, B. & Feeney, J. (2006). *J. Mol. Biol.* **356**, 886–903.

Ragona, L., Pusterla, F., Zetta, L., Monaco, H. L. & Molinari, H. (1997). *Fold. Des.* **2**, 281–290.

Raschke, T. & Marqusee, S. (1998). *Curr. Opin. Biotechnol.* **9**, 80–86.

Richardson, J. S. & Richardson, D. C. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 2754–2759.

Sali, A., Veerapandian, B., Cooper, J. B., Foundling, S. I., Hoover, D. J. & Blundell, T. L. (1989). *EMBO J.* **8**, 2179–2188.

Sali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T. & Blundell, T. L. (1992). *Proteins*, **12**, 158–170.

Schoenborn, B. P. (1969). *Nature (London)*, **224**, 143–146.

Sears, V. F. (1992). *Neutron News*, **3**, 26–37.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3872–3877.

Sukumar, N., Langan, P., Mathews, F. S., Jones, L. H., Thiyagarajan, P., Schoenborn, B. P. & Davidson, V. L. (2005). *Acta Cryst.* D**61**, 640–642.

Teeter, M. M. & Kossiakoff, A. A. (1984). *Neutrons in Biology*, edited by B. P. Schoenborn & R. B. Knott, pp. 335–348. New York: Plenum.

Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2003). *J. Am. Chem. Soc.* **125**, 15686–15687.

Wagner, G. & Wuthrich, K. (1982). *J. Mol. Biol.* **160**, 343–361.

Wlodawer, A., Borkakoti, N., Moss, D. S. & Howlin, B. (1986). *Acta Cryst.* B**42**, 379–387.

Wlodawer, A., Miller, M. & Sjolin, L. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 3628–3631.

Wlodawer, A., Savage, H. & Dodson, G. (1989). *Acta Cryst.* B**45**, 99–107.

Wlodawer, A. & Sjolin, L. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 1418–1422.

Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). *J. Mol. Biol.* **180**, 301–329.

Yamamoto, T., Izumi, S. & Gekko, K. (2004). *J. Biochem. (Tokyo)*, **135**, 17–24.

Zhang, B. & Peng, Z. (2000). *J. Mol. Biol.* **299**, 1121–1132.