HDX and Homology Modeling

HDX is often performed on known protein structures, primarily generated by X-ray crystallography and NMR, or perturbed versions of them. HDX can also be applied to systems with unknown structures to test theoretical models and generate hypotheses about function.

Homology modeling (HM) is a common approach to characterize protein structure and function especially when data from other structural approaches are not available. Although HM statistics for assigning model quality do not necessarily reflect the *in vivo* structure of proteins, the approach could be improved by adding experimental evidence that provides constraints for the model. We are testing the utility of HDX-MS results to inform HM of protein structures. HDX can reveal regions when hydrogen bond formation is affected by conformational changes. When coupled with low resolution techniques, it can provide information about the location of α-helices and β-sheets.

One example is diheme cytochrome c (DHCC), a protein for which no high resolution structure is available. It is a small protein (~24 kDa) in a family with rather simple structures that includes only helices and loops. We employed differential, solution HDX experiments at 4 °C with DHCC in buffer solution, incubated with either reductant or oxidant at a final concentration of ~25 μM. After quench and digestion, the samples can be digested by pepsin and analyzed by LC/MS (here providing 95% coverage). The extent of HDX for most regions of the protein is larger in the oxidized state, indicating that the oxidized state is more flexible (less structurally constrained). Several of the regions that display no changes upon reduction (e.g., 45-49, Figure 1) can be viewed as controls because they show no detectable exchange in either state, indicating extensive H-bonding. The largest differences in HDX kinetics occur near the heme-binding CXXCH motif (i.e., for peptides 1-29 and 109-127) (Figure 1).

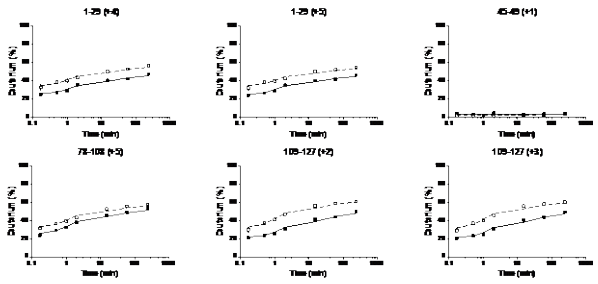

Figure 1. Kinetic curves of selected peptides used for HDX mapping for reduced (solid line) and oxidized (dash line) states of the DHCC. Numbers in parentheses with "+" sign are the charge states of the peptides.

A series of homology models were generated. First, we chose as templates 1ETP, a cytochrome c4 protein from a bacteria which might be the evolutionary origin of DHCC, and 3MK7c, which had highest alignment statistics returned by a BLAST search against the Protein Data Bank. We generated four models based on these two templates using different algorithms (Modeller, I-TASSER, and Phyre2; Figure 2). For models A and B we used 1ETP. In model A most of the regions that show discrepancies between the HDX results and the HM outcomes are the helical regions because model A does not identify the second heme binding motif.

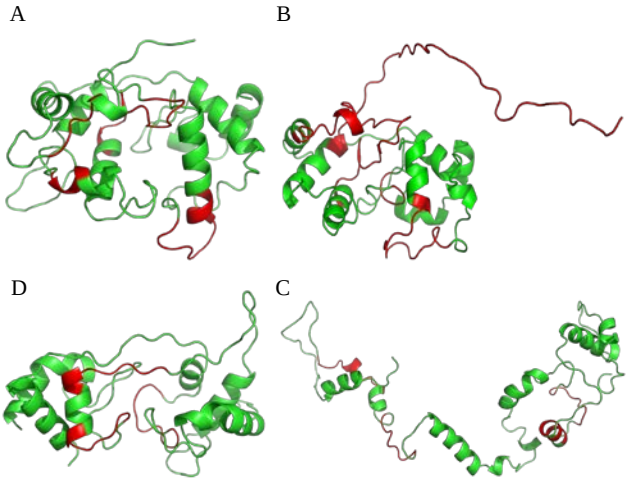

Figure 2. Homology models A to D. Regions that are highlighted in red do not agree with HDX

Model B was created by manually adjusting the sequence alignment to improve the agreement of HM and the HDX data.

Model C was developed in the same manner as models A and B, but used 3MK7c as a template. Model C gives better agreement for the second heme binding region, 109-127, than do Models A and B (Table 1). However, model C points to a very open structure in the C-terminus due to a lack of the sequence alignment. Therefore, we generated a nearly identical model D from both the online suites I-TASSER and Phyre2. Both packages identified 3MK7c as the best template, gave highly similar alignments and consequently identical models. Model D agrees best with the HDX data (Table 1 and Figure 2).

Table 1. Comparison of HDX Data and Homology Models

Region[a]	Model A[b]	Model B	Model C	Model D	HDX agreement on the models[c]			
					A	B	C	D
1-29	**h/l**	**h/l**	**h/l**	**h/l**	**	**	**	**
33-37	**h/l**	**l**	**l**	**l**	**			*
38-44	**h bundle**	**h/l**	**h/l**	**h/l**				*
45-49	**h**	**h bundle**	**h/l**	**h bundle**		**		**
50-57	**l**	**l**	**h bundle**	**h/l**				*
58-65	**l**	**l**	**l**	**l**	**	**	**	**
78-91	**h/l**	**h/l**	**h/l**	**h/l**	**	**	**	**
92-109	**h/l**	**h/l**	**h/l**	**h/l**	**	**	**	**
109-127	**l**	**l**	**h/l**	**h/l**			**	**
130-138	**h/l**	**l**	**l**	**l**	*			*
143-149	**h bundle**	**h bundle**	**h**	**h bundle**	**	**	*	**
150-168	**h/l**	**h/l**	**h/l**	**h/l**	**	**	**	**
169-172	**h**	**h bundle**	**h bundle**	**h bundle**		**	**	**
173-205	**h/l**	**l**	**h/l**	**h/l**	**	**	**	**

This demonstrates that HDX data can be used with HM to identify structural features of proteins in the absence of high resolution structural data. Compared to nuclear magnetic resonance (NMR), MS coupling with HDX has the advantage of less sample consumption, and higher upper mass limit on protein, as proteins can be digested by protease after deuterium has been incorporated. In addition, HDX-MS can work with proteins under physiologicallly relevant conditions. Proteins with impurities can also be studied directly by HDX-MS, as long as the impurities do not interfere with protein ligand interactions.