

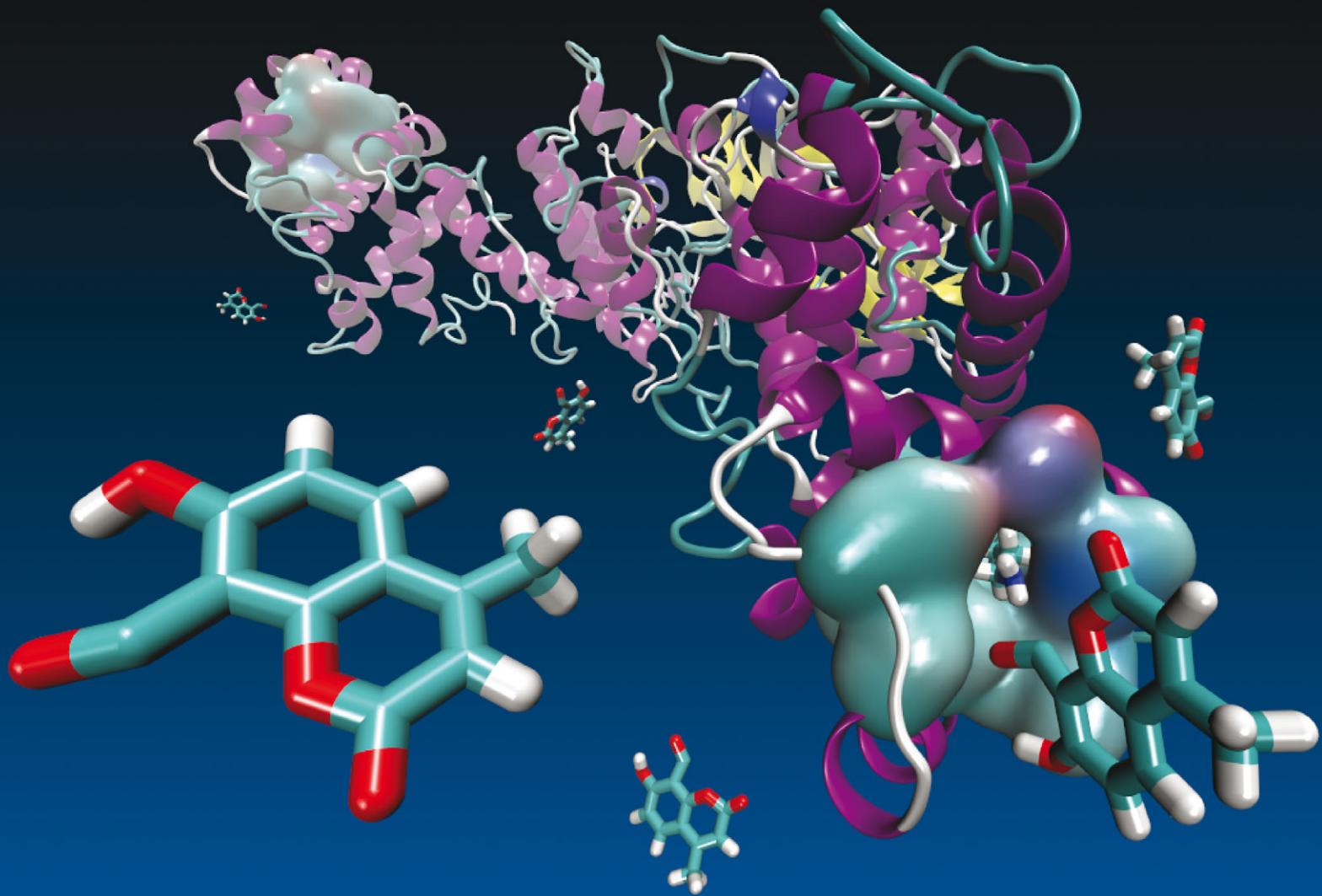
# Molecular BioSystems

Interfacing chemical biology with the -omic sciences and systems biology

[www.molecularbiosystems.org](http://www.molecularbiosystems.org)

Volume 9 | Number 10 | October 2013 | Pages 2393–2576

Indexed in  
**MEDLINE!**



ISSN 1742-206X

RSC Publishing

**REVIEW ARTICLE**

Benedict C. S. Cross, Peter J. Bond *et al.*

Selective inhibition of the unfolded protein response: targeting catalytic sites for Schiff base modification



1742-206X(2013)9:10;1-S

## Selective inhibition of the unfolded protein response: targeting catalytic sites for Schiff base modification

Cite this: *Mol. BioSyst.*, 2013, **9**, 2408

Susana M. Tomasio,<sup>a</sup> Heather P. Harding,<sup>b</sup> David Ron,<sup>b</sup> Benedict C. S. Cross<sup>\*bc</sup> and Peter J. Bond<sup>\*a</sup>

Constitutive protein misfolding in the endoplasmic reticulum (ER) can lead to cellular toxicity and disease. Consequently, the protein folding environment within the ER is highly optimised and tightly regulated by the unfolded protein response (UPR). The apparent convergence of myriad diseases upon proteostasis in the ER has triggered a broad effort to identify selective inhibitors of the UPR. In particular, the most ancient component of this cellular stress pathway, the transmembrane protein IRE1, represents an appealing target for pharmacological intervention. Several inhibitors of IRE1 have recently been reported, each containing an aldehyde moiety that forms an unusual, highly selective Schiff base with a single key lysine (K907) within the RNase domain. Here we review the progress made in chemical genetic manipulation of IRE1 and the unfolded protein response and discuss computational strategies to rationalise the selectivity of covalently active small molecules for their targets. As an exemplar, we provide additional evidence that K907 of IRE1 is buried within a particularly unusual environment that facilitates Schiff base formation. New free-energy calculations within a molecular dynamics (MD) simulation framework show that the  $pK_a$  of K907 is reduced by  $\sim 3.6$   $pK_a$  units, relative to the model  $pK_a$  of lysine in water. This significant  $pK_a$  perturbation provides additional insights into the precise requirements for inhibition and for RNase catalysis by IRE1. Our computational method may represent a general approach for identifying potential covalent inhibitory lysine sites within buried protein cavities.

Received 18th June 2013,  
Accepted 18th July 2013

DOI: 10.1039/c3mb70234k

[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

### The unfolded protein response and IRE1

The fitness and survival of eukaryotic organisms is contingent upon the fidelity of protein biosynthesis within their cells. Loss-of-function mistakes in protein folding damage cellular efficiency, whilst proteotoxic gain-of-function misfolding is associated with a heavy burden of human disease. In healthy cells, quality control mechanisms act to preclude toxicity and maintain protein folding homeostasis. However, in a diseased cell these same pathways can also be corrupted and the rectifying intentions subverted to cause dysfunction.<sup>1</sup> This apparent conflict and the convergence of multiple diseases on protein folding homeostasis has precipitated a need for new tools to understand the cellular response to perturbations in proteostasis, and garnered a broad effort for discovery.

Whilst protein biosynthesis begins in the cytosol of the cell, a large proportion of the eukaryotic proteome is trafficked through the luminal space of the endoplasmic reticulum (ER) soon after the initiation of synthesis.<sup>2</sup> Thus, the ribosome-studded rough ER is a major site of protein synthesis and this dominance is reflected in the complexity and sophistication of the systems operating to defend homeostasis in this compartment. In metazoans, accumulated unfolded or misfolded proteins are recognised by three transmembrane resident ER sensors: PERK (PKR-like endoplasmic reticulum eIF2 $\alpha$  kinase), IRE1 (inositol-requiring enzyme 1) and ATF6 (activating transcription factor 6). These components adapt the ER to the folding client load using both translational and transcriptional interventions in a pathway collectively known as the unfolded protein response (UPR). Of these components, only IRE1 is conserved from fungus to man. It is the most studied element of the UPR, is linked most promiscuously to human disease, and its unusual dual enzymatic activity is a salient draw for chemical biologists.

IRE1 exists as two isoforms in mammals: IRE1 $\alpha$  is ubiquitously expressed, whilst IRE1 $\beta$ , exhibiting only slightly differing enzymology,<sup>3</sup> is restricted to the epithelial cells of the intestinal tract. IRE1 comprises a tripartite structure with a luminal domain linked via a single transmembrane segment to the cytosolic kinase

<sup>a</sup> Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

E-mail: [pjb91@cam.ac.uk](mailto:pjb91@cam.ac.uk); Tel: +44 (0)1223 763981

<sup>b</sup> University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, UK. E-mail: [bcc33@cam.ac.uk](mailto:bcc33@cam.ac.uk); Tel: +44 (0)1223 588040

<sup>c</sup> Department of Haematology, University of Cambridge, Cambridge, CB2 2PT, UK



and endoribonuclease domains. The luminal domain of IRE1 is closely related to that of PERK and detects the accumulation of unfolded proteins by one of several proposed mechanisms. Both the direct binding of unfolded clients to IRE1<sup>4</sup> and manipulation of the membrane composition<sup>5</sup> can impact on IRE1 activity. Additionally, the constitutively IRE1-bound molecular chaperone BiP (binding immunoglobulin protein) is dislocated away from IRE1 as unfolded clients accumulate.<sup>6,7</sup> These conditions shift the monomeric pool of IRE1 towards oligomerization, triggering the kinase activity of IRE1 and trans-autophosphorylation of adjacent IRE1 protomers.<sup>8</sup> Phosphorylated IRE1 is activated, and is arranged in order to promote the endoribonuclease function that defines the role of IRE1 in the UPR. ER stress-activated IRE1 cleaves the latent mRNA for the XBP1 (X-box binding protein 1) transcription factor, liberating a 26 bp fragment and generating a frame-shifted transcript for XBP1 upon re-ligation.<sup>9,10</sup> Processed XBP1 then upregulates the expression of genes encoding ER molecular chaperones, ER biosynthetic enzymes, translocation components, ER-associated degradation factors and ER membrane biogenesis enzymes.<sup>11,12</sup> Thus, IRE1 responds to the saturation of ER client load with a broad regulatory programme that resolves ER stress by generating improved folding conditions, enhancing degradation and swelling the ER volume to thermodynamically favour folding over aggregation.<sup>13</sup> In addition to these functions, IRE1 can also cleave a subset of other ER-bound mRNAs,<sup>14</sup> in a process that may have links to cellular physiology,<sup>15,16</sup> but which at present is only partially understood. Regulated IRE1-dependent mRNA decay (RIDD) attenuates protein load on the ER by preemptively degrading ER client mRNA,<sup>17</sup> and has also recently been linked to a surprising innate immune response to infection by cholera toxin.<sup>18</sup> In this study, RIDD products were shown to act as part of a signalling cascade to initiate a cytokine response to the toxin, where IRE1 was able to directly bind the cholera toxin to trigger the effect. IRE1 can also provide a node for pro-apoptotic signals emanating from the ER *via* the ASK1 (apoptosis signalling kinase) and JNK (c-Jun N-terminal kinase) pathways.<sup>19,20</sup> This bipartisanship of IRE1 in the adaptive and apoptotic response of the cell to unfolded protein stress hints at a complex relationship with disease and places the enzyme at a compelling nexus for pharmacological intervention. Metabolic diseases including type II diabetes mellitus, neurodegeneration, cancer and inflammatory diseases have all been linked to the IRE1/XBP1 pathway,<sup>1</sup> and in most cases pathogenesis is linked with a hyperactive UPR.

## Pharmacological inhibition of the IRE1 endonuclease

The enzymology and structure of IRE1 make it an ideal candidate for reverse chemical genetics. The kinase domain of IRE1 has been targeted by nucleotide analogues, allosterically inactivating the RNase of IRE1 and blocking high order assembly of the IRE1 multimers,<sup>21</sup> but direct inhibitors of the RNase have only been discovered following large-scale screening efforts.

Remarkably, each of these studies converged upon molecules with very similar chemistry. The mammalian protein was adapted to a FRET-derepression assay for high-throughput screening and libraries of >220 000<sup>22</sup> and >240 000<sup>17</sup> compounds were screened for kinetic disruption of the fluorescence signal. In a third programme, a cell-based XBP1-luciferase assay was used to screen >60 000 compounds for their effect on *in vivo* IRE1 activity.<sup>23</sup> The lead compounds selected following each independent effort yielded small hydrophobic molecules containing an *ortho*-hydroxyl aryl aldehyde moiety: the salicylaldehydes,<sup>22</sup> the coumarin 4μ8C<sup>17</sup> and STF083010<sup>23</sup> (Fig. 1). Although the aldehyde moiety of STF083010 went undetected during the screening process, it is clear that the imine bond of the library compound rapidly hydrolyses in aqueous solution to liberate an active carbaldehyde-containing napthalene.<sup>17,24</sup> Indeed, for each of these inhibitors of IRE1, the aldehyde moiety was initially masked in the library compound by an imine linkage (Fig. 1), an important consideration when trying to deconvolute the data returned from high-throughput screening efforts and to characterise lead compounds.

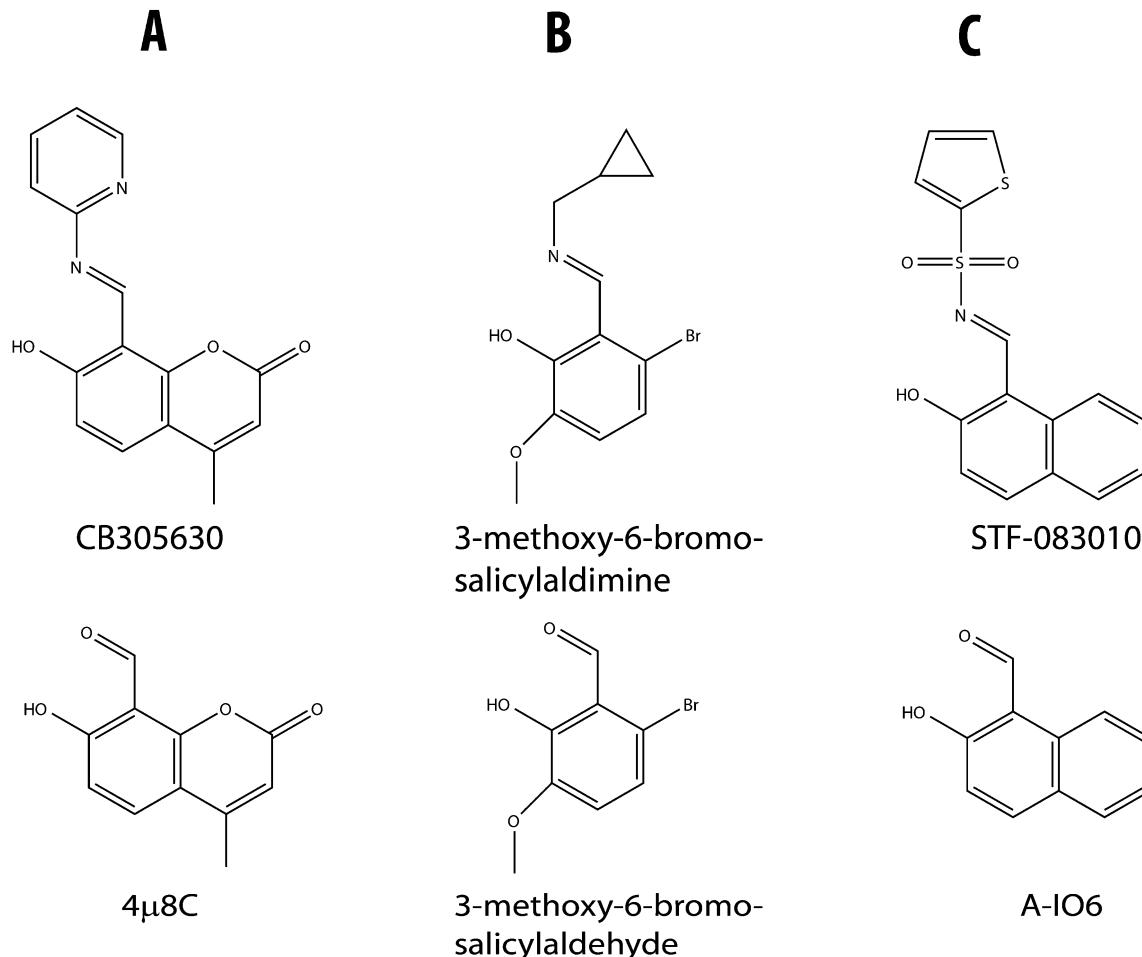
The aldehyde moiety is key to inhibition and each of these compounds operates by formation of a Schiff base with at least one key residue in the IRE1 molecule. However, despite the covalent mode of these inhibitors, they appear to retain specificity for IRE1 both *in vitro* and in cellular studies.<sup>17,22,23</sup> The explanation for the surprising selectivity of the drugs was greatly aided by biochemical analysis of 4μ8C and the application of *in silico* docking and molecular dynamics simulations to rationalise the structural implications for the drug-inhibited IRE1 molecule.

## Molecular dynamics simulations & insights into IRE1 inhibition

The molecular dynamics (MD) approach provides a Newtonian physics-based framework in which to simulate motions of biomolecules at unparalleled temporal and spatial resolution. Typically, the structure of a protein, obtained experimentally *via* nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography, is embedded in a simulation unit cell containing a physiologically modelled salt solution, resulting in a system containing tens of thousands of atoms.<sup>25,26</sup> The forces acting on each atom are calculated *via* a classical potential function or "force field" composed of simple equations containing terms for intra- and inter-molecular interactions, which may simplistically be likened to a collection of "balls and springs". Under the assumption that these forces are constant over a small time interval, they may be used to estimate accelerations, and hence new positions and velocities, of the atoms in the system, according to Newton's laws of motion.<sup>27,28</sup> Repeating this cycle many millions of times enables propagation of a simulation trajectory, from which information concerning the dynamics and energetics of biomolecules may be gathered, complementing available experimental data.

Continuing improvements in algorithms, force field parameters, and high-performance computing technologies have made it routinely possible to carry out accurate simulations of



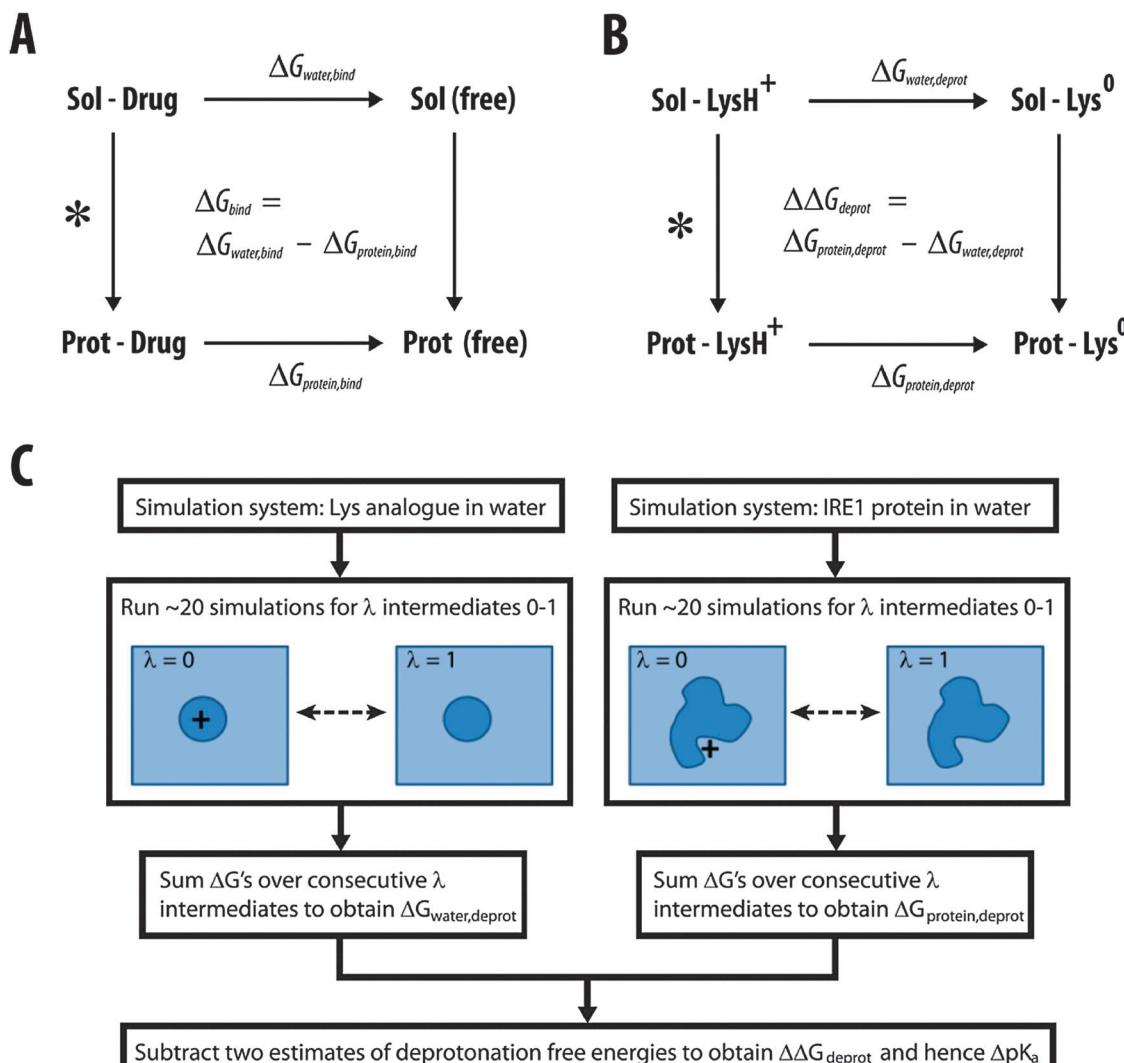


**Fig. 1** Inhibitors of the IRE1 endonuclease identified by high-throughput screening. (A) 4μ8C identified by Cross *et al.*<sup>17</sup> (B) Lead salicylaldehyde inhibitor identified by Volkman *et al.*<sup>22</sup> (C) STF-083010 identified by Papandreu *et al.*<sup>23</sup> In each of the screening programmes, the hit compounds were initially selected from imine linked scaffolds (upper panels). These were subsequently found to hydrolyse during the aqueous *in vitro* assays to yield the carbaldehyde-containing active species (lower panels).

biomolecular systems over timescales of tens or hundreds of nanoseconds, or even beyond.<sup>29</sup> MD also provides an exciting opportunity to supplement the drug design process. Proteins are dynamic and flexible; they breathe, change shape, and respond to the presence of other molecules.<sup>30</sup> Simulations can identify conformational changes in biomolecules in response to ligand binding/release,<sup>31,32</sup> and can help to discover and characterize novel druggable sites.<sup>33–35</sup> Simulations also make it possible to accurately predict thermodynamic properties such as ligand-binding free energies, *via* so-called “alchemical” transformation. In such approaches, the interaction between a drug and its environment are slowly “annihilated” during a series of MD simulations, in both the solvated, protein-bound and protein-free states.<sup>25,36,37</sup> Measuring the respective energetic changes  $\Delta G_{\text{protein,bind}}$  and  $\Delta G_{\text{water,bind}}$  during these processes enables “completion” of the thermodynamic cycle, yielding the relative free energy change for extracting a drug molecule from solvent and positioning it within the protein binding site – in other words, the binding free energy  $\Delta G_{\text{bind}}$  (Fig. 2A). Such calculations can yield extremely accurate results, but are computationally demanding, particularly as a result of their

sensitivity to inadequate conformational sampling. Thus, they are generally of most use in the drug optimization stage, following lead compound discovery.

A combination of molecular modelling, simulation, and thermodynamics calculations has recently been utilized to gain insights into the mechanism of inhibition of IRE1 by 4μ8C, complementing and extending experimental observations.<sup>17</sup> The chemical properties of 4μ8C allowed the covalent interaction of the drug with the enzyme to be directly observed by spectroscopy. This modification was further confirmed and mapped by high-performance liquid chromatography (HPLC)-linked mass spectrometry, revealing an unusually stable Schiff base at two mechanistically critical lysines. One of these (K907) is present in the endonuclease domain at the active site of IRE1 and is required for IRE1 RNase catalysis.<sup>17,38</sup> The other lysine, K599, located within the kinase domain, plays a role in phosphate coordination and is common to all kinase domains. Curiously, only K907 was modifiable *in vivo*. Thus, two key questions concerning the specificity of 4μ8C was posed by these observations: first, why does measurable Schiff-base formation by 4μ8C



**Fig. 2** Thermodynamic cycles and free-energy calculations. (A) Cycle for estimating free-energy of binding of a non-covalent drug to a protein. (B) Cycle for estimating  $pK_a$  shifts in a lysine residue on a protein site. (C) Simulation workflow for obtaining free-energy estimates to complete the thermodynamic cycle illustrated in (B). In (A) and (B),  $\Delta G_{\text{water}}$  is the free-energy change associated with the unbound drug being annihilated ( $\Delta G_{\text{water,bind}}$ ) or with an unbound model lysine sidechain becoming deprotonated ( $\Delta G_{\text{water,deprot}}$ ).  $\Delta G_{\text{protein}}$  is the equivalent free-energy change in the protein-bound state for annihilation ( $\Delta G_{\text{protein,bind}}$ ) or deprotonation ( $\Delta G_{\text{protein,deprot}}$ ), respectively. In (A), measurement of these quantities enables calculation of the free-energy  $\Delta G_{\text{bind}}$  for drug binding along the \* arm. In (B), the \* arm represents  $\Delta\Delta G_{\text{deprot}}$ , the relative free-energy difference for deprotonation between the protein-resident and solvated forms of lysine, which enables calculation of the  $pK_a$  shift within the protein relative to the model compound, as illustrated in (C).

only occur at two out of a total of twenty-five lysine residues in its cytosolic domains? Second, why is the modifiability of K599 context-dependent?

To begin to answer these questions, Cross *et al.* performed explicitly solvated, all-atom MD simulations of the “apo” uninhibited state of the IRE1 dimer.<sup>17</sup> The complete system amounted to a size on the order of a quarter of a million atoms, making these calculations computationally demanding. The first hint to the peculiar specificity of 4 $\mu$ 8C came from an analysis of the exposed surface area accessible to solvent of the lysine residues present in IRE1. Most lysines exhibited an average exposed surface area of  $\sim 90 \text{ \AA}^2$  during simulation on the nanosecond timescale. This agrees reasonably well with a previous analysis of protein structures from the protein databank

(PDB), showing that lysine is the most solvent-accessible of all amino acids, with a mean exposed area of  $\sim 100 \text{ \AA}^2$ .<sup>39</sup> With the exception of two residues buried at the dimer interface, the only lysines in IRE1 that did not follow this trend were K599 and K907, each with exposed areas of just  $\sim 50 \text{ \AA}^2$ . Only  $\sim 10\%$  of lysines in folded proteins tend to exhibit this degree of burial, and for comparison, fully exposed lysine possesses a surface area of over  $200 \text{ \AA}^2$ .<sup>39</sup>

Thus, whilst most lysines in the cytosolic domain of IRE1 are surface exposed and flexible, K599 and K907 remain somewhat shielded from bulk solvent. Following Schiff base formation by 4 $\mu$ 8C at these sites, reduced ease of hydrolysis of the imine bond compared to other sites was hypothesized to lead to a low off-rate, helping to explain the selectivity. On the other hand,

this did not provide an explanation for the lack of modifiability of K599 *in vivo*. To answer this, *in silico* docking of 4μ8C, along with extensive geometry optimization, of the inhibitor-bound state in the presence and absence of ADP-Mg<sup>2+</sup> was performed. Strikingly, favourable, unstrained binding of 4μ8C was only possible in the absence of nucleotide; in other words, the two were predicted to sterically compete for the same site. This mutually exclusive binding is consistent with experimental evidence where co-incubation of the inhibitor with nucleotide prevents the binding of 4μ8C to IRE1.<sup>17</sup> Likewise, in the context of the whole cell, binding of 4μ8C to IRE1 was prevented by endogenous competing nucleotide.<sup>17</sup>

## Predicted mode of IRE1 inhibition

To gain further insights into the structural basis for IRE1 inhibition, possible 4μ8C-inhibition modes were modelled, using a tailored protocol designed for the covalently-bound inhibitor involving extensive sampling and optimization in torsional-angle space around each rotatable bond of the Schiff base at K907.<sup>17</sup> A number of possible orientations for 4μ8C were identified *via* this approach, the most favourable (large and negative enthalpy of interaction) of which satisfied all possible interactions within the binding site. As shown in Fig. 3A, this involves a stacked conformation of one of the 4μ8C coumarin rings against the sidechain of F889, along with two hydrogen-bonds to the sidechain of D885 and/or H910. The predicted location of 4μ8C would be expected to interfere with catalysis and impede substrate binding, by forming a wedge positioned between the catalytic residues Y892 and H910.

We now extend these observations to the other families of small, potent inhibitory molecules of IRE1: a hydrolysed carbaldehyde-containing naphthalene analogue of STF-083010<sup>23</sup> and one of the most potent salicylaldehydes, 3-methoxy-6-bromo-salicylaldehyde.<sup>22</sup> The same exhaustive conformational sampling and optimization protocol was used to identify favourably bound states of each to K907, as described previously.<sup>17</sup> In each case, a similar interaction mode was predicted for the most favourable orientation as for 4μ8C (Fig. 3B and C). For the STF-083010 analogue, an almost identical conformation was

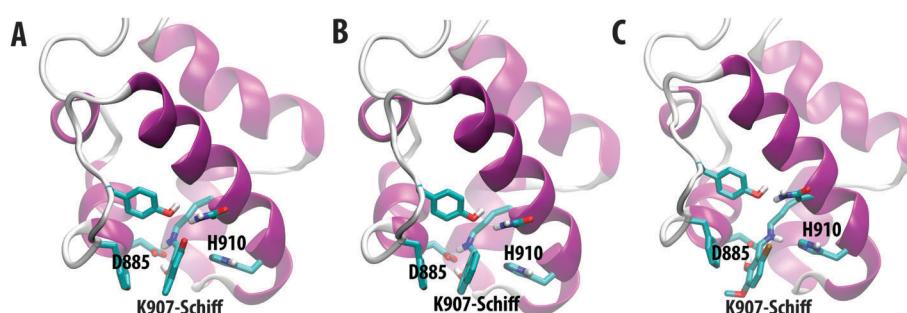
observed, with comparable stacking of one of the rings against F889, and two hydrogen-bonds to D885. For the salicylaldehyde, a hydrogen-bond was formed with H910 to facilitate stacking of the inhibitor's single ring against F889, again leading to a similar inhibitory mode within the catalytic site. Thus, these independently identified aldehyde-containing small molecules appear to have converged upon a similar mechanism of inhibition.

To test the stability of this predicted binding mode, the modelled conformation of 4μ8C bound to K907 was used to initiate multiple MD simulations of the inhibited IRE1 in solvent.<sup>17</sup> In each subsequent simulation replica, the 4μ8C rings remained stacked against the sidechain of F889, and the inhibitor relaxed into a hydrophobic pocket formed by sidechains of nearby, nonpolar residues. Strikingly, this immersion of 4μ8C within the hydrophobic pocket appeared to constrain the RNase active site, which by comparison, was relatively dynamic over tens of nanoseconds in the apo, unbound state. This would be expected to further slow the off-rate by reducing bulk water exchange and hence hydrolysis at the imine bond site.

## Computational determination of IRE1 K907 pK<sub>a</sub>

The reduced solvent exposure of the K907 sidechain and the stable burial of bound 4μ8C within a hydrophobic microenvironment are of great interest, since conditions of low polarity or polarizability can shift pK<sub>a</sub> values of internal ionizable residues including lysines<sup>40</sup> towards their unbound form. The ε-amino group of K907 must be in its uncharged form in the first step of Schiff base formation, to enable nucleophilic attack on the 4μ8C carbaldehyde carbonyl. We therefore questioned whether the hydrophobic microenvironment accessible to K907, buried within the RNase active site, might down-shift its pK<sub>a</sub>, and thus increase the rate of Schiff base formation by 4μ8C. To investigate this further, we note that the dissociation constant K<sub>a</sub> of a titratable group is related to the free energy change of deprotonation  $\Delta G_{\text{deprot}}$  in the following way:

$$\Delta G_{\text{deprot}} = -kT \ln K_a, \text{ and thus } K_a = e^{-\frac{\Delta G_{\text{deprot}}}{kT}}$$



**Fig. 3** IRE1 inhibition modes predicted for (A) 4μ8C,<sup>17</sup> (B) naphthalene analogue of STF-083010,<sup>23</sup> and (C) 3-methoxy-6-bromo-salicylaldehyde.<sup>22</sup> The RNase domain is shown in cartoons format, with key residues highlighted in wireframe format and labelled. The binding conformation for each inhibitor was predicted *via* sampling and optimization in torsional-angle space around each Schiff base rotatable bond, using the CHARMM package<sup>55</sup> with the CHARMM22/CMAP<sup>41</sup> and CGenFF<sup>56</sup> forcefields for protein and inhibitor, respectively. Molecular graphics were generated using VMD.<sup>57</sup>



where  $k$  is the Boltzmann constant. The  $pK_a$  may therefore be related to  $\Delta G_{\text{deprot}}$ :

$$pK_a = -\log_{10} K_a = \frac{\Delta G_{\text{deprot}}}{2.303kT}$$

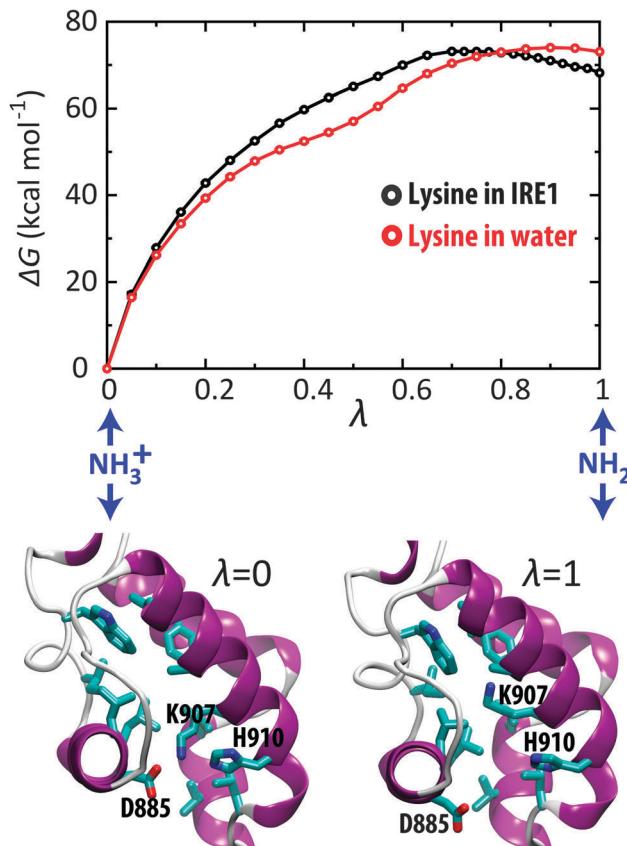
As discussed above, alchemical transformation within an MD simulation framework may be used to estimate free-energy changes associated with, for example, repositioning a solvated drug molecule into a protein binding site, as part of a thermodynamic cycle (Fig. 2A). In the same way, the free-energy associated with deprotonation may be calculated, *via* annihilation of the proton associated with an ionizable site. Thus, according to Fig. 2B, calculation of the free energy for deprotonation of an ionisable residue in the protein ( $\Delta G_{\text{protein,deprot}}$ ) and for the model amino acid analogue in solvent ( $\Delta G_{\text{water,deprot}}$ ) enables estimation of a  $pK_a$  shift,  $\Delta pK_a$ :

$$\Delta pK_a = \frac{\Delta \Delta G_{\text{deprot}}}{2.303kT},$$

where  $\Delta \Delta G_{\text{deprot}} = \Delta G_{\text{protein,deprot}} - \Delta G_{\text{water,deprot}}$

Two sets of free-energy calculations were therefore performed, of the complete IRE1 dimer in water (to estimate  $\Delta G_{\text{protein,deprot}}$ ), and of a lysine amino acid analogue in water (to estimate  $\Delta G_{\text{water,deprot}}$ ). In each case, 20–30 MD simulations of 3–4 ns each were performed, in which the system was gradually changed *via* a coupling parameter  $\lambda$ , from state  $\lambda = 0$  (K907 protonated) to state  $\lambda = 1$  (K907 deprotonated). The respective free energy changes (averaged over both IRE1 monomers) could then be calculated as a sum of energies over the consecutive, intermediate states. Each simulation was carried out using the same conditions as described previously.<sup>17</sup> Briefly, the system contained dimeric IRE1 (PDB ID code 3P23) in complex with ADP-Mg<sup>2+</sup>, surrounded by a 0.1 M NaCl solution in a truncated octahedral box ( $\sim 250\,000$  atoms total), and was described using the CHARMM22/CMAP<sup>41</sup> forcefield within the GROMACS 4.5 simulation package.<sup>42,43</sup>

As shown in Fig. 4, the cumulative free energy change for deprotonation of isolated lysine in water ( $\Delta G_{\text{water,deprot}} = 73.1 \pm 0.6$  kcal mol<sup>-1</sup>) and of K907 in IRE1 ( $\Delta G_{\text{protein,deprot}} = 68.2 \pm 0.6$  kcal mol<sup>-1</sup>) yields  $\Delta \Delta G_{\text{deprot}} = -4.9 \pm 0.8$  kcal mol<sup>-1</sup>. We therefore estimate a  $pK_a$  down-shift for K907 within IRE1 of  $3.6 \pm 0.6$  units. Thus, relative to the model  $pK_a$  of lysine in water ( $\sim 10.5$ ), the  $pK_a$  of the K907 sidechain is likely to lie at around 7, and macroscopically, it will exist in IRE1 molecules as an equal population of charged and uncharged states at neutral pH. At the structural level (Fig. 4), this  $pK_a$  shift results from the ease with which K907 becomes buried within the deep hydrophobic pocket upon deprotonation, combined with the dynamic nature of the RNase active site entrance, as observed previously.<sup>17</sup> For example, upon complete deprotonation, the  $\epsilon$ -amino group of K907 moves further into the binding site, and its mean distance from D885 and H910 at the active site entrance increases, respectively, from  $0.5 \pm 0.2$  Å and  $0.2 \pm 0.1$  Å during simulation at  $\lambda = 0$ , to  $0.7 \pm 0.2$  Å and  $0.5 \pm 0.1$  Å during simulation at  $\lambda = 1$ .

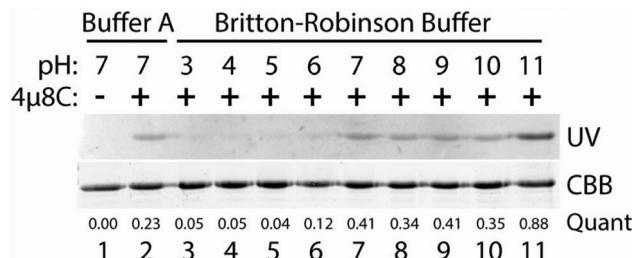


**Fig. 4** Calculated  $pK_a$  shift for K907 in IRE1. Above, free-energy curves are shown for K907 in IRE1 ( $\Delta G_{\text{protein}}$ ) and for a lysine sidechain in water ( $\Delta G_{\text{model}}$ ), with corresponding snapshots shown below of the simulated structures of the RNase domain in its protonated ( $\lambda = 0$ ) and deprotonated ( $\lambda = 1$ ) forms. The RNase domain is shown in cartoons format. Hydrophobic residues within the binding pocket, along with key ionizable sidechains, are shown in wireframe format. Molecular graphics were generated using VMD.<sup>57</sup> The free-energy curves were generated using the Bennett acceptance ratio (BAR) method with soft-core potentials.  $\Delta G_{\text{protein}}$  was calculated using  $\lambda$  windows 0.05 or 0.025 apart, between  $0 \leq \lambda < 0.7$  and  $0.7 \leq \lambda \leq 1.0$  respectively.  $\Delta G_{\text{model}}$  used  $\lambda$  windows 0.05 apart between  $0 \leq \lambda \leq 1.0$ . Extensive energy minimisation followed by 0.2 ns equilibration was performed, prior to the production run for each window. Each simulation was run under conditions of constant temperature (298 K)<sup>58</sup> and pressure (1 atm),<sup>59</sup> using a 2 fs time step with LINCS constraints applied to bond lengths.<sup>60</sup> Electrostatics were calculated using the Particle-Mesh Ewald algorithm, and Van der Waals interactions were smoothly switched off from 10 to 12 Å.

Experimental support for this shift in  $pK_a$  at lysine 907 is found when the efficiency of 4 $\mu$ 8C modification of IRE1 is measured *in vitro*. Binding of the fluorescent compound to IRE1 can be visualised in-gel following UV irradiation,<sup>17</sup> and we find that whilst this modification is strictly pH-dependent, reactivity is observed even in conditions where lysines on IRE1 are expected to be deprotonated (Fig. 5). Substantially increased binding of 4 $\mu$ 8C is found when the buffered conditions match the  $pK_a$  for lysine in solution where the compound is able to react with the other nucleophiles on IRE1 ( $>\text{pH } 10$ , Fig. 5, lane 11). Hence, the  $pK_a$  perturbed K907 is able to participate in Schiff base formation even at pH 7.

Overall, our results suggest that K907 is more easily deprotonated than usual under normal cell conditions, consistent with





**Fig. 5** pH-dependent binding of 4μ8C to IRE1. Compound was incubated with the human IRE1 cytosolic domain<sup>17</sup> at the indicated pH before addition of sodium borohydride and analysis by electrophoresis. Bound 4μ8C was visualised by UV excitation and the protein was then stained with coomassie brilliant blue (CBB). Quantification lane shows the calculated relative fraction of the protein that was found modified under each condition.

it being buried in a hydrophobic microenvironment that is partially dehydrated. This would lead to increased nucleophilicity of the K907 ε-amino group, an increased rate of Schiff base formation with 4μ8C, and a slow off-rate.

## RNase catalysis in IRE1

In addition to the contribution to the Schiff base formation, a shifted  $pK_a$  at K907 may have consequences for the mechanism of RNase catalysis by IRE1. Structural analysis of the yeast apoenzyme provided the first detailed clues into the mechanism of endoribonucleolytic scission in IRE1.<sup>44</sup> Catalysis by IRE1 is likely to proceed *via* a general acid–base mechanism, similar to that found for *Archaeoglobus fulgidus* splicing endonuclease and yielding a 2',3'-cyclic phosphate product.<sup>44,45</sup> His1061/Tyr1043 in yeast IRE1 would therefore form a proton transfer acid–base pair for scission, homologous to His257/Tyr246 of the splicing endonuclease and analogous to His pairs in both RNase A and T1. R1056 in yeast IRE1 stabilises the transition state by hydrogen bond formation with the non-bridging oxygen during proton transfer to enact RNA cleavage.<sup>46,47</sup> The contribution to catalysis for the two core catalytic residues in ScIre1p has been quantitatively demonstrated,<sup>46</sup> and elements of the active site are also well conserved to mammalian IRE1 and RNase L. This core mechanism therefore likely largely persists through evolution, although some notable divergence in the active site from the yeast to the human enzyme is apparent. In the structure for the human enzyme for example,<sup>48</sup> the conserved HsIRE1α R902 is found displaced from the core catalytic pair whilst the sidechain of K907 is oriented towards the His/Tyr intersection. Thus, in mammalian IRE1, K907 appears to be better placed to support transition state stabilisation than R902, and given the pronounced reduction of the  $pK_a$  it is likely to perform this role better, forming a stronger hydrogen bond to the cleavage intermediate. A similar arrangement is found in RNase A where K41 is  $pK_a$  perturbed, lowering the free energy of the transition state to promote catalysis.<sup>49</sup> Our computationally derived model may therefore suggest a mechanistic rationale for excelled catalysis by human IRE1 as compared to the budding yeast enzyme – a more potent mammalian enzyme provides the

necessary depth to cleavage for the emergence of RIDD and the physiological consequences of lapsed specificity. Interestingly, although the *S. cerevisiae* IRE1 does not appear to conduct a RIDD-like process, the fission yeast enzyme does.<sup>50</sup> Solution of the *S. Pombe* IRE1 structure might therefore provide important new clues to the evolution of IRE1 cellular function.

## Outlook

Large-scale screening efforts have led to the independent discovery of several hydrophobic inhibitors that target the IRE1 RNase domain. Each molecule contains a common aldehyde moiety, whose presence enables selective Schiff base formation with K907.<sup>17,22,23</sup> A combination of biochemical analysis, molecular modelling, and simulation has provided a rationale for this mode of inhibition, and in particular, immersion of each molecule within the hydrophobic pocket of the RNase domain appears to contribute to specificity. As well as slowing the off-rate by protecting the imine bond against hydrolysis, we now believe that this binding mode also directly influences the chemistry of inhibition. The K907 sidechain must be in its uncharged form in the first step of Schiff base formation, to facilitate nucleophilic attack on the inhibitor. The surrounding, nonpolar microenvironment has indeed been shown to perturb the  $pK_a$  of K907 to around 7. This would favour deprotonation and increase its intrinsic nucleophilicity, making it significantly reactive even at neutral pH, as supported by our experimental data. This is not unprecedented. A lysine within an aldolase catalytic antibody was previously demonstrated to proceed *via* a Schiff base intermediate; its deeply buried location within a hydrophobic pocket perturbed its  $pK_a$  below 7.<sup>51</sup> Moreover, such shifts can be functionally important in catalysis, and indeed, measurements based upon engineered staphylococcal nuclease constructs have revealed down-shifts in internal lysine residues by as much as 5  $pK_a$  units.<sup>52</sup> In this context, our results also hint at the source of differing catalytic efficiencies in human and yeast IRE1. Uncovering the divergence in enzymology in the active site of IRE1 by both computational and biochemical means may therefore provide important new understanding of these cellular pathways.

As well as providing a useful new tool for manipulating IRE1 *in vivo* and *in vitro*, these newly discovered classes of inhibitors provide a possible route to treatments for diseases associated with proteostasis. For many years, covalent drugs have stimulated anxiety in the pharmaceutical industry due to their potential for off-target reactivity. Nevertheless, several of the top-selling drugs are covalent inhibitors of their targets (e.g. proton pump inhibitors), leading to suggestions that we are likely to see a resurgence of interest in covalent drugs.<sup>53</sup> Whilst existing covalent drugs have tended to be discovered through biological screening assays, it seems possible that a more targeted approach may now be possible. As suggested by Singh *et al.*, design could involve identifying non-reversible inhibitors for a specific target site, followed by adaptation of the structure for covalent binding.<sup>53</sup> Given the emerging perspective that  $pK_a$ -shifted lysines in buried protein cavities may be more common than expected,<sup>54</sup>



we could therefore envisage a general drug design pipeline which incorporates the identification of nearby lysine residues, combined with simulation-based  $pK_a$  shift calculations, towards selective Schiff-base inhibition.

## Acknowledgements

We wish to thank the Darwin Supercomputer of the University of Cambridge and the Swiss National Supercomputing Center *via* DECI/PRACE-2IP for computational resources. PJB and SMT thank Unilever and EPSRC for funding. HPH, DR and BCSC were supported by a Wellcome Trust Principal Research Fellowship to DR.

## References

- 1 S. Wang and R. J. Kaufman, *J. Cell Biol.*, 2012, **197**, 857–867.
- 2 B. C. S. Cross, I. Sinning, J. Luirink and S. High, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 255–264.
- 3 T. Iwawaki, A. Hosoda, T. Okuda, Y. Kamigori, C. Nomura-Furuwatari, Y. Kimata, A. Tsuru and K. Kohno, *Nat. Cell Biol.*, 2001, **3**, 158–164.
- 4 B. M. Gardner, D. Pincus, K. Gotthardt, C. M. Gallagher and P. Walter, *Cold Spring Harbor Perspect. Biol.*, 2013, **5**, a013169.
- 5 R. Volmer, K. van der Ploeg and D. Ron, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 4628–4633.
- 6 A. Bertolotti, Y. Zhang, L. M. Hendershot, H. P. Harding and D. Ron, *Nat. Cell Biol.*, 2000, **2**, 326–332.
- 7 D. Pincus, M. W. Chevalier, T. Aragon, E. van Anken, S. E. Vidal, H. El-Samad and P. Walter, *PLoS Biol.*, 2011, **8**, e1000415.
- 8 C. E. Shamu and P. Walter, *EMBO J.*, 1996, **15**, 3028–3039.
- 9 M. Calfon, H. Zeng, F. Urano, J. H. Till, S. R. Hubbard, H. P. Harding, S. G. Clark and D. Ron, *Nature*, 2002, **415**, 92–96.
- 10 H. Yoshida, T. Matsui, A. Yamamoto, T. Okada and K. Mori, *Cell*, 2001, **107**, 881–891.
- 11 A. H. Lee, N. N. Iwakoshi and L. H. Glimcher, *Mol. Cell Biol.*, 2003, **23**, 7448–7459.
- 12 K. J. Travers, C. K. Patil, L. Wodicka, D. J. Lockhart, J. S. Weissman and P. Walter, *Cell*, 2000, **101**, 249–258.
- 13 J. E. Chambers, K. Petrova, G. Tomba, M. Vendruscolo and D. Ron, *J. Cell Biol.*, 2012, **198**, 371–385.
- 14 J. Hollien, J. H. Lin, H. Li, N. Stevens, P. Walter and J. S. Weissman, *J. Cell Biol.*, 2009, **186**, 323–331.
- 15 D. Han, A. G. Lerner, L. Vande Walle, J. P. Upton, W. Xu, A. Hagen, B. J. Backes, S. A. Oakes and F. R. Papa, *Cell*, 2009, **138**, 562–575.
- 16 K. Y. Hur, J. S. So, V. Ruda, M. Frank-Kamenetsky, K. Fitzgerald, V. Koteliansky, T. Iwawaki, L. H. Glimcher and A. H. Lee, *J. Exp. Med.*, 2012, **209**, 307–318.
- 17 B. C. S. Cross, P. J. Bond, P. G. Sadowski, B. K. Jha, J. Zak, J. M. Goodman, R. H. Silverman, T. A. Neubert, I. R. Baxendale, D. Ron and H. P. Harding, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E869–E878.
- 18 J. A. Cho, A. H. Lee, B. Platzer, B. C. Cross, B. M. Gardner, H. De Luca, P. Luong, H. P. Harding, L. H. Glimcher, P. Walter, E. Fiebiger, D. Ron, J. C. Kagan and W. I. Lencer, *Cell Host Microbe*, 2013, **13**, 558–569.
- 19 H. Nishitoh, A. Matsuzawa, K. Tobiume, K. Saegusa, K. Takeda, K. Inoue, S. Hori, A. Kakizuka and H. Ichijo, *Genes Dev.*, 2002, **16**, 1345–1355.
- 20 F. Urano, X. Wang, A. Bertolotti, Y. Zhang, P. Chung, H. P. Harding and D. Ron, *Science*, 2000, **287**, 664–666.
- 21 L. Wang, B. G. Perera, S. B. Hari, B. Bhattacharai, B. J. Backes, M. A. Seeliger, S. C. Schurer, S. A. Oakes, F. R. Papa and D. J. Maly, *Nat. Chem. Biol.*, 2012, **8**, 982–989.
- 22 K. Volkmann, J. L. Lucas, D. Vuga, X. Wang, D. Brumm, C. Stiles, D. Kriebel, A. Der-Sarkissian, K. Krishnan, C. Schweitzer, Z. Liu, U. M. Malyankar, D. Chiovitti, M. Canny, D. Durocher, F. Sicheri and J. B. Patterson, *J. Biol. Chem.*, 2011, **286**, 12743–12755.
- 23 I. Papandreu, N. C. Denko, M. Olson, H. Van Melckebeke, S. Lust, A. Tam, D. E. Solow-Cordero, D. M. Bouley, F. Offner, M. Niwa and A. C. Koong, *Blood*, 2011, **117**, 1311–1314.
- 24 C. L. Kriss, J. A. Pinilla-Ibarz, A. W. Mailloux, J. J. Powers, C. H. Tang, C. W. Kang, N. Zanesi, P. K. Epling-Burnette, E. M. Sotomayor, C. M. Croce, J. R. Del Valle and C. C. Hu, *Blood*, 2012, **120**, 1027–1038.
- 25 J. D. Durrant and J. A. McCammon, *BMC Biol.*, 2011, **9**, 71.
- 26 R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu and D. E. Shaw, *Annu. Rev. Biophys.*, 2012, **41**, 429–452.
- 27 S. Khalid and P. J. Bond, *Methods Mol. Biol.*, 2013, **924**, 635–657.
- 28 T. Paramo, D. Garzon, D. A. Holdbrook, S. Khalid and P. J. Bond, *Methods Mol. Biol.*, 2013, **974**, 435–455.
- 29 D. W. Borhani and D. E. Shaw, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 15–26.
- 30 D. L. Mobley and K. A. Dill, *Structure*, 2009, **17**, 489–498.
- 31 D. Garzon, C. Anselmi, P. J. Bond and J. D. Faraldo-Gomez, *J. Biol. Chem.*, 2013, **288**, 19528–19536.
- 32 P. J. Bond and J. D. Faraldo-Gomez, *J. Biol. Chem.*, 2011, **286**, 25872–25881.
- 33 R. Baron and J. A. McCammon, *Annu. Rev. Phys. Chem.*, 2013, **64**, 151–175.
- 34 C. J. Woods, M. Malaisree, N. Pattarapongdilok, P. Sompornpisut, S. Hannongbua and A. J. Mulholland, *Biochemistry*, 2012, **51**, 4364–4375.
- 35 M. J. Harvey and G. De Fabritiis, *Drug Discovery Today*, 2012, **17**, 1059–1062.
- 36 M. R. Shirts, *Methods Mol. Biol.*, 2012, **819**, 425–467.
- 37 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
- 38 W. Tirasophon, K. Lee, B. Callaghan, A. Welihinda and R. J. Kaufman, *Genes Dev.*, 2000, **14**, 2725–2736.
- 39 L. Lins, A. Thomas and R. Brasseur, *Protein Sci.*, 2003, **12**, 1406–1417.
- 40 T. K. Harris and G. J. Turner, *IUBMB Life*, 2002, **53**, 85–98.
- 41 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prothom, C. Simmerling and J. A. Wolfson, *J. Phys. Chem. B*, 2004, **108**, 11909–11923.



W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

42 B. Hess, C. Kutzner, D. Van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

43 P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess and E. Lindahl, *J. Chem. Theory Comput.*, 2010, **6**, 459–466.

44 K. P. Lee, M. Dey, D. Nucleai, C. Cao, T. E. Dever and F. Sicheri, *Cell*, 2008, **132**, 89–100.

45 T. N. Gonzalez, C. Sidrauski, S. Dorfler and P. Walter, *EMBO J.*, 1999, **18**, 3119–3132.

46 A. V. Korennykh, A. A. Korostelev, P. F. Egea, J. Finer-Moore, R. M. Stroud, C. Zhang, K. M. Shokat and P. Walter, *BMC Biol.*, 2011, **9**, 47.

47 S. Xue, K. Calvin and H. Li, *Science*, 2006, **312**, 906–910.

48 M. M. Ali, T. Bagratuni, E. L. Davenport, P. R. Nowak, M. C. Silva-Santisteban, A. Hardcastle, C. McAndrews, M. G. Rowlands, G. J. Morgan, W. Aherne, I. Collins, F. E. Davies and L. H. Pearl, *EMBO J.*, 2011, **30**, 894–905.

49 J. M. Messmore, D. N. Fuchs and R. T. Raines, *J. Am. Chem. Soc.*, 1995, **117**, 8057–8060.

50 P. Kimmig, M. Diaz, J. Zheng, C. C. Williams, A. Lang, T. Aragon, H. Li and P. Walter, *eLife*, 2012, **1**, e00048.

51 C. F. Barbas, 3rd, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Bjornestedt, B. List, J. Anderson, E. A. Stura, I. A. Wilson and R. A. Lerner, *Science*, 1997, **278**, 2085–2092.

52 W. E. Stites, A. G. Gittis, E. E. Lattman and D. Shortle, *J. Mol. Biol.*, 1991, **221**, 7–14.

53 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discovery*, 2011, **10**, 307–317.

54 D. G. Isom, C. A. Castaneda, B. R. Cannon and B. Garcia-Moreno, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 5260–5265.

55 B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

56 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, Jr., *J. Comput. Chem.*, 2010, **31**, 671–690.

57 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38, 27–38.

58 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

59 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

60 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

