



CrossMark  
 click for updates

Cite this: *RSC Adv.*, 2015, 5, 86303

# Cross-strand disulfides in the non-hydrogen bonding site of antiparallel $\beta$ -sheet (aCSDns): poised for biological switching†

Naomi L. Haworth<sup>ab</sup> and Merridee A. Wouters<sup>\*cd</sup>

Forbidden disulfides are stressed disulfides found in recognisable protein contexts previously defined as structurally forbidden. The torsional strain of forbidden disulfides is typically higher than for structural disulfides, but not so high as to render them immediately susceptible to reduction under physiological conditions. The meta-stability of forbidden disulfides makes them likely candidates as redox switches. Here we mined the Protein Data Bank for examples of the most common forbidden disulfide, the aCSDn. This is a canonical motif in which disulfide-bonded cysteine residues are positioned directly opposite each other on adjacent anti-parallel  $\beta$ -strands such that the backbone hydrogen bonded moieties are directed away from each other. We grouped these aCSDns into homologous clusters and performed an extensive physicochemical and informatic analysis of the examples found. We estimated their torsional energies using quantum chemical calculations and studied differences between the preferred conformations of the computational model and disulfides found in solved protein structures to understand the interaction between the forces imposed by the disulfide linkage and typical constraints of the surrounding  $\beta$ -sheet. In particular, we assessed the twisting, shearing and buckling of aCSDn-containing  $\beta$ -sheets, as well as the structural and energetic relaxation when hydrogen bonds in the motif are broken. We show the strong preference of aCSDns for the right-handed staple conformation likely arises from its compatibility with the twist, shear and  $C_{\alpha}$  separation of canonical  $\beta$ -sheet. The disulfide can be accommodated with minimal distortion of the sheet, with almost all the strain present as torsional strain within the disulfide itself. For each aCSDn cluster, we summarise the structural and strain data, taxonomic conservation and any evidence of redox activity. aCSDns are known substrates of thioredoxin-like enzymes. This, together with their meta-stability, means they are ideally suited to biological switching roles and are likely to play important roles in the molecular pathways of oxidative stress.

Received 5th June 2015  
 Accepted 24th September 2015

DOI: 10.1039/c5ra10672a

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

## Introduction

Historically, disulfide bonds have generally been considered to act primarily as stabilisers of protein structure. Early work which suggested an alternate role for disulfides, as redox-dependent switches of protein function, did not gain wide currency. A 1980 review by Buchanan highlighted the importance of redox switching of protein disulfide bonds in the control of energy metabolism in chloroplasts under ambient light.<sup>1</sup> Although Buchanan noted that disulfide-based redox regulation was also important in contexts other than photosynthesis, this insight only recently gained prominence when it

was recognised that similar regulatory systems are found in all forms of life on earth.<sup>2,3</sup> This has led to the paradigm of two disulfide proteomes: one involved in structural stabilization and the other in redox control.<sup>4</sup> Thiol-based redox control is now established as an important cellular signalling system that likely arose prior to oxygenation of the earth's atmosphere, and thus predates phosphorylation.<sup>3</sup> Molecular pathways of thiol-based redox control are compromised in diseases of oxidative stress, including cancer, neurodegenerative diseases, ageing and heart disease.<sup>5</sup> Better understanding of these pathways will undoubtedly lead to improved therapeutics for these diseases, helping to relieve their huge socio-economic burden.

Redox-active disulfides have physicochemical characteristics which are distinct from those of structurally-stabilising disulfides. The electronic environment<sup>6</sup> and physical stresses within the protein<sup>7</sup> can lower the stability of a disulfide, allowing it to be more easily reduced. Physical stress can be induced dynamically due to protein conformational changes. Alternatively, it can be an inherent property of a protein. Of particular interest is physical stress resulting from disulfides located in

<sup>a</sup>Life and Environmental Sciences, Deakin University, Geelong 3217, Vic, Australia

<sup>b</sup>Victor Chang Cardiac Research Institute, Sydney 2010, NSW, Australia

<sup>c</sup>Olivia Newton-John Cancer Research Institute, Heidelberg 3084, Vic, Australia.  
 E-mail: [merridee.wouters@onjcri.org.au](mailto:merridee.wouters@onjcri.org.au)

<sup>d</sup>School of Cancer Medicine, La Trobe University, Melbourne, 3086, Vic, Australia

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5ra10672a



primary or secondary structure contexts which require either the protein structure to be distorted, or the disulfide to be twisted into a strained conformation, to allow the two cysteine (Cys) residues to be linked. In 1981, Richardson and Thornton developed a series of rules (hereafter referred to as the RT rules) describing “forbidden” protein contexts; that is, contexts in which the strain required to bring the sulfur atoms into the close proximity required for oxidation was expected to be too high to allow disulfide bonds to form.<sup>8–10</sup> However, we have recently identified disulfides occupying each of these forbidden contexts.<sup>11</sup> By far the largest group of these “Forbidden Disulfides” (FDs) break RT rule A – that is, they connect two cysteine residues located on adjacent strands of  $\beta$ -sheet (or  $\beta$ -ribbon) structures.<sup>11</sup> We refer to these as “Between Strand Disulfides” (BSDs). Significantly, in several cases BSDs have been shown to perform functional roles in their resident proteins. In particular, it has been shown that the redox states of some BSDs change in response to external stimuli (such as changes in the redox potential of the milieu or docking of a cofactor), allowing the disulfides to perform important roles in biological functions such as catalysis and cell signalling.<sup>12–16</sup>

Several different BSD types are found amongst the structures of the Protein Data Bank.<sup>11</sup> The most common and best recognised of these is the Cross Strand Disulfide (CSD). CSDs connect two Cys residues which are immediately opposite each other in the  $\beta$  sheet, *i.e.* linking residues *i* and *j* on adjacent  $\beta$ -strands. There are three different structural contexts in regular  $\beta$ -sheet in which the two Cys residues can oxidise to give a CSD motif.<sup>11</sup> For CSDs in antiparallel  $\beta$ -sheet (aCSDs), the two Cys can either be disposed across a non-hydrogen bonding (NHB) site, giving rise to an aCSDn motif; or across a hydrogen bonding (HB) site, resulting in an aCSDh. CSDs are also found in parallel  $\beta$ -sheet (pCSDs). (See Fig. S1.†)

Here we investigate in detail disulfides which adopt the aCSDn structural motif. This is by far the most common BSD motif found in the PDB.<sup>11</sup> It was first demonstrated to be a stable structure in  $\beta$ -peptides in 1978,<sup>17</sup> before also being found in protein structures in the 1990s.<sup>18–21</sup> The number of aCSDns in protein structures in the PDB is increasing relative to other disulfides, likely due to the solution of more challenging structures that contain redox-active disulfides as technical advances in X-ray crystallography are made.<sup>22</sup> Several groups have investigated the properties of aCSDns found in the PDB,<sup>23–26</sup> focussing on the conformations adopted by the disulfides and the associated torsional strain. These investigations have consistently found that aCSDns have an extremely strong preference to adopt a right-handed staple disulfide conformation. In fact, it has been stated that this is the only conformation which allows the Cys  $C_{\alpha}$  atoms to come into the close proximity required for cross-strand  $\beta$ -partners.<sup>27</sup> However, the analogous left-handed staple conformation should produce identical  $C_{\alpha}$ – $C_{\alpha}$  distances. Despite this, it is very rarely seen in protein structures. Here we investigate this handedness preference.

Earlier works predominantly focussed on true-aCSDns, that is: aCSDns which are longitudinally embedded in the middle of  $\beta$  structures (Fig. 1A and B).<sup>19,21,23–25</sup> In this study we describe

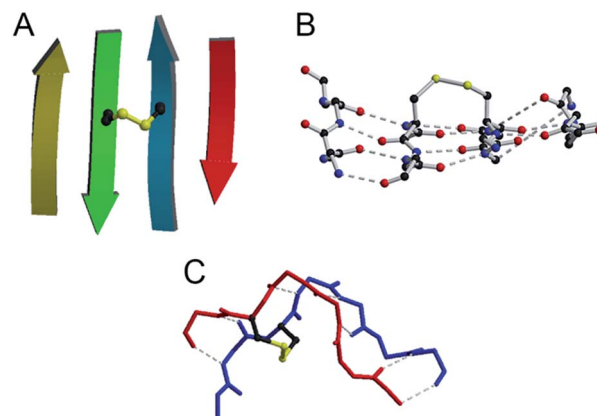


Fig. 1 (A and B) A representative disulfide in an aCSDn motif (true-aCSDn-2) from human amyloid serum P component: PDB 2a3y A Cys 36 to Cys 95. Note: the introduction of the disulfide does not cause noticeable buckling of the  $\beta$ -sheet (only the usual sheet twist is observed) and the hydrogen bonds are locally flat. (C) A representative subclass 3 true-aCSDn (disulfide immediately adjacent to an A-bent bulge) from the surface active protein hydrophobin: PDB 1r2m A Cys 14 to Cys 26. Figure prepared using Molscript<sup>28</sup> and Raster3D.<sup>29</sup>

other subclasses of aCSDns, as well as examining more closely the interactions between aCSDns and their surrounding  $\beta$ -sheets. The analysis will focus on local factors such as the disulfide conformation and the resulting torsional strain on the disulfide bond, as well as the strain exerted by the aCSDn on the surrounding  $\beta$ -sheet, *i.e.* the backbone conformations enforced on the involved Cys residues, and the twisting and shearing of the sheet due to the presence of the aCSDn. The relative populations of each aCSDn subclass within a normalized PDB dataset and their biological roles will be discussed throughout. Some of this information has been touched on briefly in previous work,<sup>11</sup> however here we provide full details and analysis as well as examples and exceptions. Subsequent works<sup>30,31</sup> will provide a similar analysis for the aCSDh and pCSD BSD motifs.

## Results and discussion

### A aCSDn structures

(a) **Sub-motifs and subclasses.** The nature of the hydrogen bonding surrounding a BSD is a critical determinant of its physical and chemical properties, and its potential biological activity.<sup>11</sup> For this reason we earlier highlighted the importance of distinguishing between BSD motifs – that is, whether or not the Cys residues are involved in cross-strand hydrogen bonding (aCSDns *vs.* aCSDhs). We also developed concepts of disulfide embeddedness. The degree to which a disulfide is embedded within  $\beta$ -structure determines both the level of strain in the system due to the presence of the disulfide and its accessibility to reducing agents. Longitudinal embeddedness (LonE) measures the degree of canonical hydrogen bonding on either side of each Cys residue within the disulfide-containing  $\beta$ -ladder, whereas lateral embeddedness (LatE) describes the presence of additional  $\beta$ -chains hydrogen bonded to either



side of the  $\beta$ -ladder in which the disulfide is found. There are three possible levels of LonE for aCSDNs, giving three different aCSDn sub-motifs: true aCSDNs, where the disulfide is embedded in the middle of a  $\beta$ -sheet; end-aCSDNs where it is at the end of a  $\beta$ -ladder; and  $\beta$ -bridge-aCSDNs where the disulfide connects two residues which form an isolated  $\beta$ -bridge structure. Hydrogen bonding diagrams of these three sub-motifs are shown in Fig. 2. In total, 4862 aCSDNs were identified in our dataset of protein structures from the PDB. The distribution of these aCSDNs between the three possible sub-motifs is shown in Fig. 3.

The true-aCSDn is the most common BSD sub-motif found in our PDB dataset, being represented by 3232 individual protein structures. As some proteins are overrepresented in the PDB, a more accurate description of the relative abundance of each BSD sub-motif in the PDB snapshot can be obtained by collating disulfides together into homologous "clusters".<sup>11</sup> 157 unique clusters were found where the disulfide always adopts a true-aCSDn sub-motif. Examples of true-aCSDNs with known functions include: the catalytic disulfide between Cys 103 and Cys 109 in the N-terminal domain of DsbD (PDB: 116p);<sup>12</sup> a regulatory disulfide between Cys 130 and Cys 159 in CD4 that is reduced upon entry of HIV into the cell (PDB: 3o2d);<sup>13,32</sup> and a true-aCSDn between Cys 186 and Cys 209 in tissue factor (PDB:

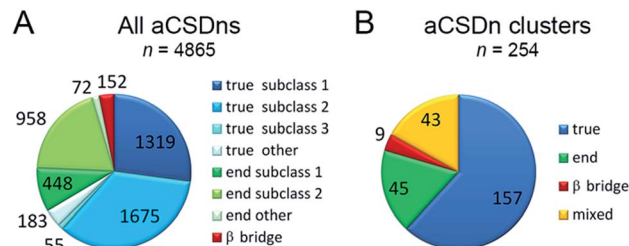


Fig. 3 (A) The distribution of the aCSDNs in our dataset amongst the three possible sub-motifs and their subclasses. (B) The distribution of non-homologous aCSDn clusters in our dataset amongst the sub-motifs. The "mixed" group includes all clusters where there is heterogeneity in the sub-motifs observed.

2c4f) that switches the protein between its coagulation and cell signalling functions.<sup>14</sup>

We have identified three major subclasses of true-aCSDNs amongst the structures of the PDB. These subclasses are distinguished by their hydrogen bonding patterns. In the first group, both pairs of residues flanking the disulfide-bonded Cys form two hydrogen bonds; and both Cys, as well as all four of their neighbouring residues, adopt  $\beta$  backbone conformations (subclass 1, 1319 structures). In the second subclass of true-aCSDNs one of the outermost hydrogen bonds is missing (subclass 2, 1675 structures). In some cases this allows the flanking residue(s) on that side of the disulfide to adopt backbone conformations in the  $\alpha$  region of the Ramachandran plot.<sup>33</sup> It is not uncommon for true-aCSDNs belonging to a particular disulfide cluster to be split between subclasses 1 and 2. This indicates that the variable hydrogen bond is very weak in this protein and may be broken in some structures. It is potentially also an indication of conformational flexibility in the neighbourhood of the disulfide. For instance, in roughly three quarters of the instances of bovine  $\beta$ -lactoglobulin in our dataset, the disulfide connecting Cys 106 to Cys 119 belongs to subclass 1 (e.g. PDB: 1beb); while in the remainder the motif corresponds to subclass 2 (e.g. PDB: 2blg). The comparatively high B-factors for the residues in the Cys–Cys loop for many  $\beta$ -lactoglobulin structures supports the notion of conformational flexibility. For example, in PDB: 2blg the average B-factor for residues 107 to 118 is  $72.6 \pm 23.7$  vs. the overall average of  $50.1 \pm 23.4$ . In a third, much rarer, subclass of true-aCSDNs, the disulfide is located immediately adjacent to an A-bent bulge (Fig. 1C, 55 structures, 10 clusters). In these structures the two residues on the bulged side of the disulfide occupy the  $\alpha$  region of the Ramachandran plot.<sup>33</sup> Examples of this subclass are found in hydrophobin (e.g. PDB: 1r2m Cys 14 to Cys 26), thaumatin-like proteins (PDB: 2pe7 Cys 51 to Cys 61) and FKBP13, which is involved in the immune response in plants and animals (PDB: 1u79 Cys 106 to Cys 111). In FKBP13 the redox state of the disulfide is known to regulate the isomerase activity of the protein.<sup>34</sup> There are also 183 true-aCSDNs which cannot be allocated to any of the three subclasses described above. In the vast majority of cases, their structures are of lower resolution ( $>2.0 \text{ \AA}$ ) and the disulfides belong to clusters where most instances are either subclass 1 or 2 true-aCSDNs. Other

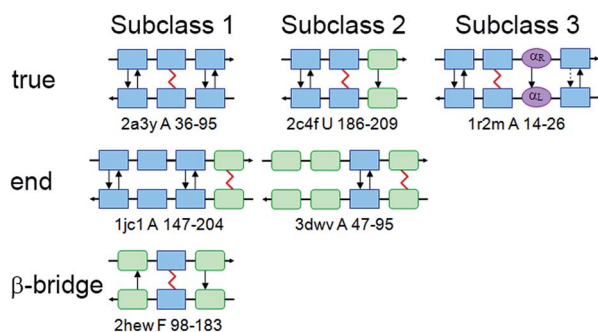


Fig. 2 Hydrogen bonding diagrams for aCSDn sub-motifs (rows) and subclasses (columns) identified in this study. In each case an example PDB structure from our dataset is given. Three subclasses of the true-aCSDn sub-motif are shown: 1, true-aCSDn fully embedded in regular  $\beta$  structure; 2, true-aCSDn with an adjacent residue pair forming only one hydrogen bond; 3, true-aCSDn immediately adjacent to an A-bent bulge. There is structural evidence to suggest that subclasses 1 and 2 may be interconvertible. Two end-aCSDn subclasses are also illustrated: 1, end-aCSDn at the end of a  $\beta$ -ribbon; 2, end-aCSDn adjacent to an HB site  $\beta$ -bridge. Residues which consistently adopt a  $\beta$  backbone conformation are shown as blue rectangles and those which are always found in  $\alpha$  conformation are purple ovals. Any residue which can adopt either backbone conformation is shown as a green rectangle with rounded corners. Hydrogen bonds are shown as arrows. Dashed arrows mean that the hydrogen bond may be missing in some structures. The disulfide bond is depicted as a red zig-zag connecting two Cys residues. The examples shown are: serum amyloid P component (PDB: 2a3y), coagulation factor VIIa (PDB: 2c4f), the surface active protein hydrophobin (PDB: 1r2m), an engineered redox-sensitive green fluorescent protein (PDB: 1jc1), tryparedoxin peroxidase (PDB: 3dvw) and tumour necrosis factor 4 ligand (DPB: 2hew).



instances of atypical aCSDns may be examples of rarer subclasses, for example aCSDns located immediately adjacent to classic  $\beta$ -bulges, or enclosing  $\beta$ -hairpin loops with five or less residues between the two Cys.

True-aCSDns are found with all three possible levels of lateral embeddedness:  $\beta$ -ribbon-BSDs or BSD-0s, which have no  $\beta$ -chains hydrogen bonded to the  $\beta$ -ladder containing the disulfide; edge-BSDs or BSD-1s with one adjacent chain; and mid-sheet-BSDs or BSD-2s which have adjacent chains on both sides of the disulfide.<sup>11</sup>

A large population of aCSDns (1478 structures) are also found at the termini of adjacent  $\beta$ -strands. Although there are roughly half as many instances of these “end”-aCSDns in our database as there are true-aCSDns, many of the end-aCSDn structures are redundant; for example, over 50% are found in eukaryotic trypsin-like serine protease (TLSP) structures (linking Cys 136 and Cys 201). There are only 45 clusters in which the disulfide is consistently an end-aCSDn.

In antiparallel sheet, end-aCSDns occur outside the last hydrogen bonded residue pair of the sheet (subclass 1, 448 structures) or linking two residues immediately adjacent to an HB site  $\beta$ -bridge (subclass 2, 958 structures). In both subclasses the lack of hydrogen bonding on one side of the disulfide allows the Cys residues to adopt non- $\beta$  backbone conformations. In most cases, however, both Cys do occupy the  $\beta$  region of the Ramachandran plot. An exception is the end-aCSDn found in the catalytic domain of glycosyl hydrolase family 12 (e.g. PDB: 2nlr Cys 5 to Cys 31), where one of the involved Cys consistently adopts an  $\alpha$  conformation. This does not appear to affect the disulfide bond, see Fig. S2.† Other examples of end-aCSDn sub-motifs are the disulfide connecting Cys 126 to Cys 196 in the HIV/SIV cell entry protein gp120 (e.g. PDB: 2nxy) which has been shown to be both redox active<sup>35</sup> and essential for virus infectivity;<sup>36</sup> and the catalytic disulfide (Cys 47 to Cys 95) of euglenozoan glutathione peroxidases (PDB: 3dww).<sup>37</sup> As for true-aCSDns, a small fraction of end-aCSDns (72) cannot be allocated to either subclass 1 or 2. Again, these are mostly low-resolution structures, or are associated with  $\beta$ -hairpin loops where there is no hydrogen bonding on the loop side of the disulfide (e.g. procathepsin X, PDB: 1deu Cys 112 to Cys 118). All three possible levels of LatE are seen for disulfides in end-aCSDn sub-motifs.

aCSDns are also found within NHB site  $\beta$ -bridge structures (see Fig. 2, row 3). These are much rarer than true- and end-aCSDns; only 152 disulfides in our dataset adopt this sub-motif. There were only nine clusters in which the disulfides were exclusively  $\beta$ -bridge-aCSDns and only two additional clusters in which a significant fraction of the disulfides adopted this sub-motif. Examples of  $\beta$ -bridge-aCSDns are found in OX40L (a TNF ligand superfamily member, PDB: 2hew Cys 98 to 183), in cellobiohydrolase 1 (glycosyl hydrolase family 7, e.g. PDB: 1ojk Cys 215 to Cys 234) and in the serine carboxypeptidase family of proteins (e.g. PDB: 1ivy Cys 60 to Cys 334). In all cases the backbone  $\phi$  and  $\psi$  angles of both Cys residues occupy the core  $\beta$  region of the Ramachandran plot. As for true- and end-aCSDn sub-motifs,  $\beta$ -bridge-aCSDns are seen with all three

levels of LatE. Hence, LatE is largely independent of the disulfide sub-motif or subclass.

#### (b) Variations of sub-motif within disulfide clusters.

Several clusters within our database contain homologous disulfides belonging to different aCSDn sub-motifs. These are designated as ‘mixed’ in Fig. 3. Some cases may be due to dynamic variation in the protein secondary structure, while others may arise from evolutionary variation between homologous proteins.

In some cases, heterogeneity in the disulfide sub-motif may be an indication of weak hydrogen bonding in the vicinity of the disulfide. An example is the aCSDn connecting Cys 822 and Cys 884 of the calf-2 domain of integrin  $\alpha$ . This disulfide lies on the edge of a  $\beta$ -sheet. Hydrogen bonding between the internal and edge strands is detected on both sides of the disulfide in the A chain of PDB: 3fcs (true-aCSDn) but only on one side (end-aCSDn) in the C chain of the same structure. This disulfide has been shown to be redox-sensitive.<sup>38</sup> It has been proposed that reduction of the aCSDn allows the edge strand to shift in relation to its neighbour, resulting in switching between the “bent” (inactive) and “upright” (active) conformations of integrin  $\alpha$ .<sup>38</sup> This will be discussed in more detail below.

Short, flexible sections of protein chain near a disulfide (such as  $\beta$ -hairpins or small loops between the two Cys) can adopt multiple conformations, some of which may allow a true-aCSDn sub-motif to form, while others may permit only an end-aCSDn to be present.<sup>39</sup> One such example involves a disulfide connecting Cys 436 and Cys 445 in *Clostridium botulinum* neurotoxin (e.g. PDB: 1epw vs. PDB: 1f31). Reduction of this disulfide and snipping of the intermediate  $\beta$ -hairpin from the polypeptide chain results in formation of a protein-conducting channel which transmits the catalytic domain of the toxin through cell membranes.<sup>40</sup> A second example is found on an exposed  $\beta$ -hairpin loop of scytalidopepsin B and other peptidase A4 proteins (e.g. PDB: 2ifr Cys 141 to Cys 148 vs. PDB: 2ifw). Binding of the substrate peptide by this  $\beta$ -hairpin results in a change in conformation of the loop and, in some cases, a change in the aCSDn sub-motif.<sup>41</sup>

Finally, heterogeneity in the aCSDn sub-motif adopted can be due to evolutionary variation. Some disulfides are conserved between proteins of different species (orthologs), proteins deriving from duplications (paralogs), or even amongst several protein families within a SCOP superfamily.<sup>11</sup> In these cases, differences in primary or secondary structure between proteins may result in different levels of LonE. For instance, in  $\beta$ -lactamase inhibitory protein (BLIP, PDB: 3gmu) the disulfide connecting Cys 109 to Cys 131 is a true-aCSDn. However, in a BLIP-like protein (BLP, PDB: 3gmx) a proline residue is adjacent to one of the Cys; precluding cross-strand hydrogen bond formation, and resulting in an end-aCSDn sub-motif.

(c) **Interdependence of LonE and LatE.** As noted earlier, the longitudinal and lateral embeddedness of a disulfide affect its inherent strain and its accessibility to physiological reducing agents, such as thioredoxin and glutathione. Although all possible combinations of lateral and longitudinal embeddedness are observed in our aCSDn dataset, when comparing



disulfide clusters, indications of the interdependence of LonE and LatE begin to emerge.

For instance,  $\beta$ -bridge-aCSDns tend to be found in  $\beta$ -ribbons. When in edge-of-sheet and mid-sheet LatE contexts, they are almost always seen in clusters where the disulfide is usually associated with a true-aCSDn sub-motif. These are also often found in low resolution structures or near strained hydrogen bonds (for example, neighbouring a  $\beta$ -bulge). This suggests that embedded  $\beta$ -bridge-aCSDn sub-motifs are unstable transient states, which may possibly precede reduction.

In contrast, clusters which adopt true- and end-aCSDn sub-motifs are more evenly spread between the different LatE levels. Edge-of-sheet LatE contexts are the most common, being seen in roughly half of all true- and end-aCSDn clusters. This is

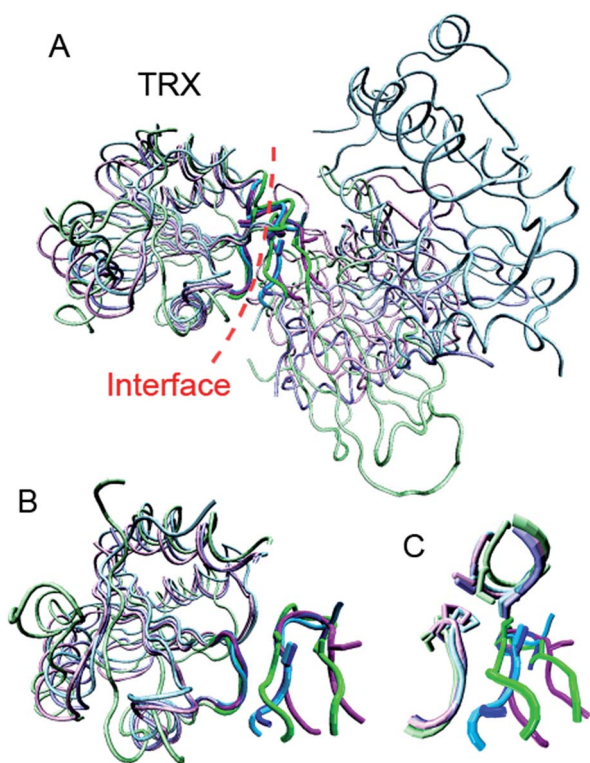
consistent with the expectation that many BSDs are likely to preferentially interact with thioredoxin-like (Trx-like) enzymes.<sup>11</sup> Thioredoxin binds to an edge strand of its substrate protein in order to correctly position the target disulfide for reduction by its catalytic CXXC motif (Fig. 4). For example, the transmembrane electron transporter, DsbD, contains an edge-of-sheet aCSDn which is known to interact with Trx-like proteins (PDB: 2k9f, green in Fig. 4). In nature, DsbD activates oxidized periplasmic DsbC by reducing the Trx-like DsbC active site Cys using electrons transported from the cytoplasm. In the final step of the electron transport process, the true-aCSDn-1 connecting Cys 103 and Cys 109 in the N-terminal domain of DsbD is reduced by its Trx-like C-terminal domain.<sup>12</sup> The interaction of Trx with aCSDns will be discussed in further detail below.

For clusters which are dominated by true-aCSDn sub-motifs,  $\beta$ -ribbon and mid-sheet LatE contexts are seen in roughly half as many clusters as edge-of-sheet LatE (~30%  $\beta$ -ribbon, ~60% edge-of-sheet, ~30% mid-sheet). The distribution is less balanced for clusters dominated by end-aCSDn sub-motifs, with  $\beta$ -ribbon and edge-of-sheet LatE contexts each being seen in ~50% of the clusters, and only ~10% being BSD-2s. The latter group, however, contains the eukaryotic TLSP cluster; this means that ~50% of the individual end-aCSDn structures are found in the middle of  $\beta$ -sheets. In summary, aCSDns are most commonly found in edge-of-sheet contexts, consistent with the expectation that they are thioredoxin substrates.

As for LonE, structures in some clusters are split between two or more LatE levels (hence the percentages in the previous paragraph adding up to more than 100%). Such heterogeneity may be biologically significant, indicating that the aCSDn is involved in a substrate binding site or the site of a protein-protein interaction. Examples include: the disulfide in the peptide binding site of peptidase A4 (mentioned earlier), which is found on a  $\beta$ -ribbon in the free enzyme (PDB: 1s2k) and on an edge strand when binding the substrate peptide (PDB: 2ifr); a true-aCSDn found on an edge strand in Bowman-Birk inhibitors (e.g. PDB: 2g81, I chain, Cys 51 to Cys 59) which can bind trypsin along its free edge (e.g. the homologous disulfide connecting Cys 24 to Cys 32 of PDB: 2g81 I); and a disulfide in the binding surface of the pilus chaperone caf1m (e.g. PDB: 1z9s vs. PDB: 2os7, Cys 98 to Cys 137) which has been proposed to be involved in redox processes.<sup>43</sup> Care must be taken in this analysis as sometimes the observed heterogeneity in LatE is a result of incomplete crystal structures (missing domains, unresolved residues, etc.). For instance, an aCSDn in the HIV and SIV gp120 envelope glycoprotein (Cys 218 to Cys 247) appears to be on a  $\beta$ -ribbon in PDB: 2nxy, but on an edge strand in PDB: 3jwd. Closer inspection reveals that the N-terminus, containing the additional  $\beta$ -strand, is not resolved in the former structure.

## B Disulfide conformations and stability

(a) **Conformations observed for aCSDns.** As described in our earlier work,<sup>11</sup> the stability of a disulfide is dependent on its conformation. The conformation is defined by the three central dihedral angles of the disulfide bond:  $\chi_2$  ( $C_\alpha C_\beta S_\gamma S'_\gamma$ ),  $\chi_3$



**Fig. 4** (A) Crystal structures of thioredoxin interacting with several non-homologous substrates. Thioredoxin structures, shown on the left, superimpose well, while the substrate proteins on the right do not. Green: *Neisseria meningitidis* thioredoxin interacting with DsbD, PDB: 2k9f. Blue: *Saccharomyces cerevisiae* thioredoxin 2 interacting with peptide methionine sulfoxide reductase, PDB: 3pin. Magenta: *Hordeum vulgare* thioredoxin H interacting with barley  $\alpha$ -amylase/subtilisin inhibitor (BASI), PDB: 2iwt. Cyan: *Escherichia coli* thioredoxin 1 interacting with phosphoadenosine phosphosulfate reductase, PDB: 2o8v. Structures superimposed using the Magic Fit function of Swiss-PDBViewer.<sup>42</sup> (B) Detail from part (A) highlighting the binding loop of thioredoxin (left) specifically recognising aCSDns of the substrates (right). The CXXC motif of thioredoxin has been removed for clarity. Although the substrates are not homologous, their target aCSDns superimpose well. (C) Detail from part (A) showing the interacting portions of thioredoxin (pastel) and its substrates. Here, the catalytic CXXC motif of thioredoxin is included (top) along with the disulfides of the substrates (right). The *cis*-proline residue of the binding loop of thioredoxin is also shown.



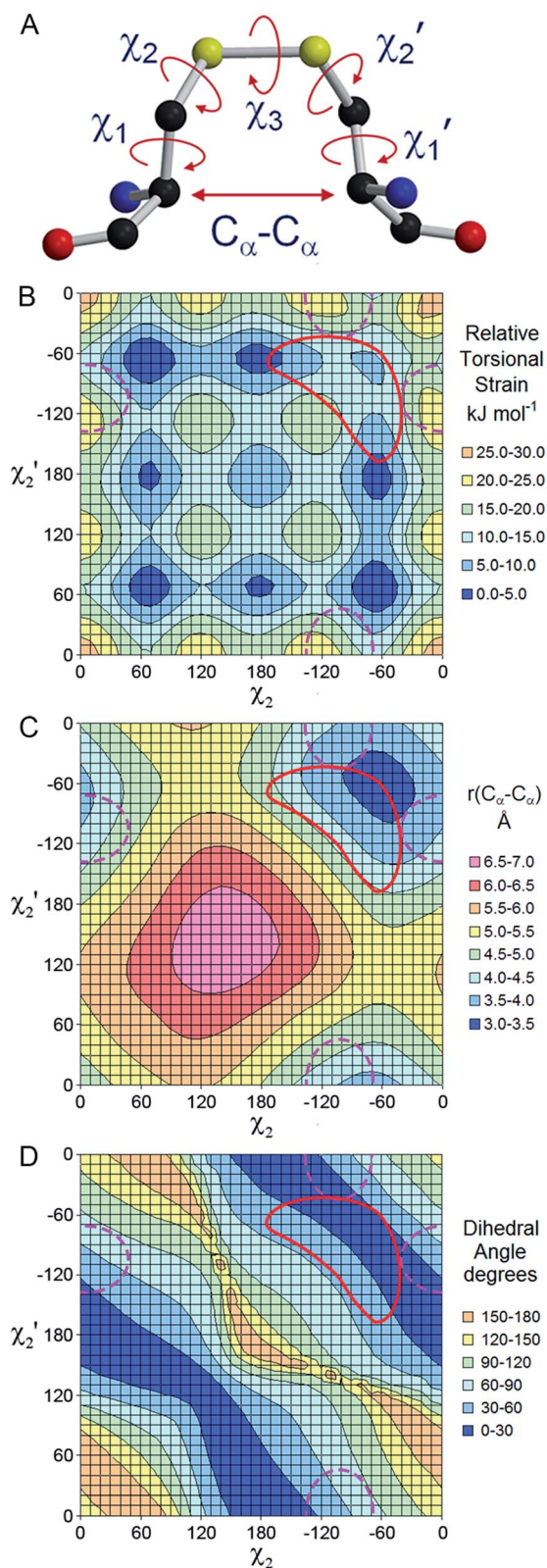


Fig. 5 (A) A representative aCSDn structure defining the dihedral angles  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_2'$  and  $\chi_1'$  and the  $C_\alpha-C_\alpha$  distance. (B) Contour plot of the variation in the torsional strain in the disulfide bond (in kJ mol<sup>-1</sup>) with torsion around  $\chi_2$  and  $\chi_2'$  ( $\chi_3 = 100^\circ$ , the mean value for aCSDns). Blue-er regions are lower in energy and red-er regions are more highly strained. Blue and cyan regions represent conformations with torsional strain less than 15 kJ mol<sup>-1</sup> – any disulfides outside these regions

( $C_\beta S_\gamma S_\gamma' C_\beta'$ ) and  $\chi_2'$  ( $S_\gamma S_\gamma' C_\beta' C_\alpha'$ ), see Fig. 5A. Twisting of these dihedral angles away from their optimal values produces torsional strain in the disulfide bond. We refer to the strain resulting from torsion of these three dihedral angles as the disulfide torsional energy (DTE).<sup>11</sup> The  $\chi_1$  and  $\chi_1'$  dihedral angles of the two involved Cys can also be strained. We refer to the sum of the DTE and the strain in  $\chi_1$  &  $\chi_1'$  as the total torsional energy (TTE). Although disulfides can adopt any low energy conformation,<sup>11</sup> only a restricted selection of conformations are compatible with an aCSDn motif.

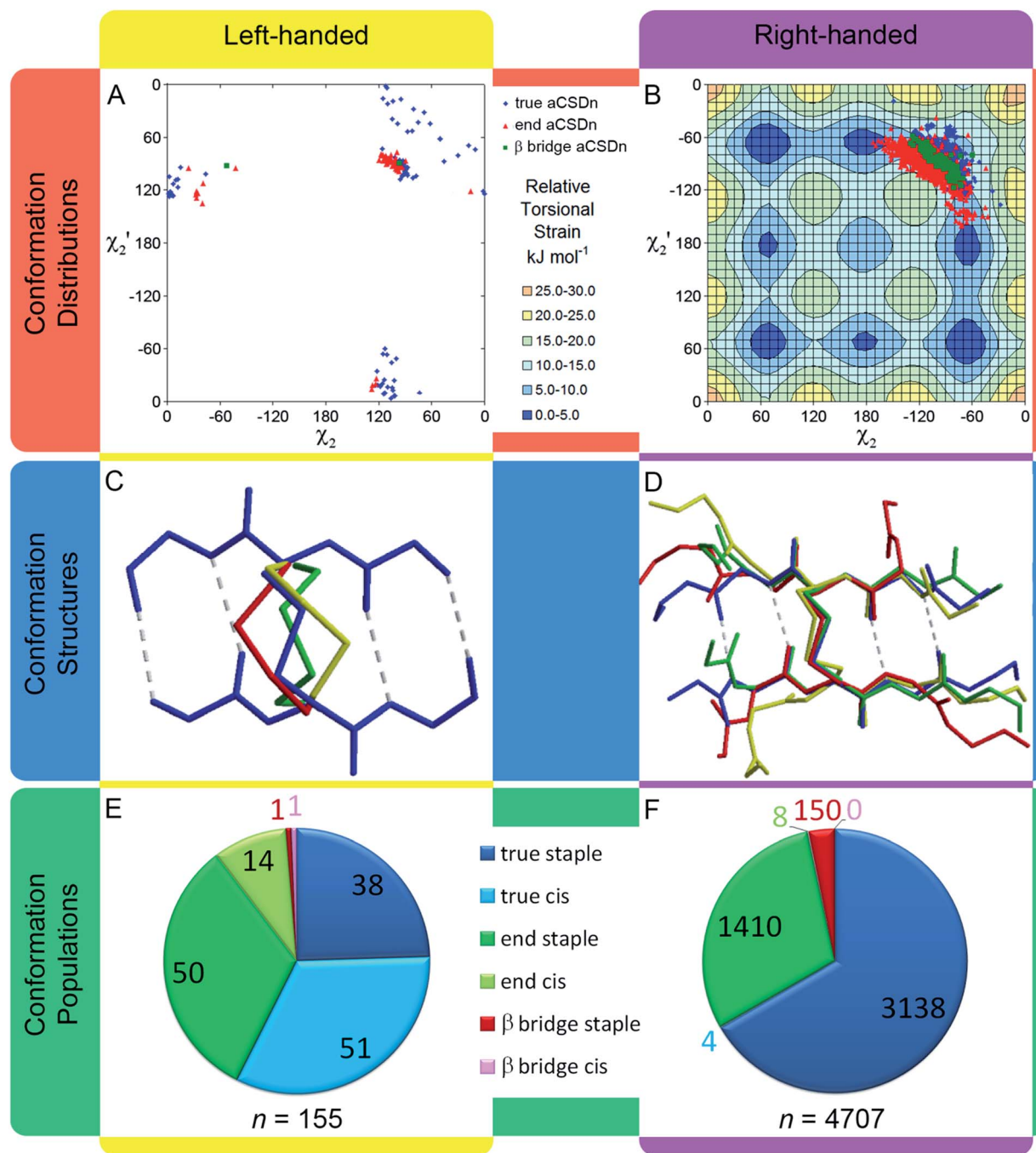
There have been many reports that aCSDns highly favour right-handed staple disulfide conformations: that is, conformations with the  $\chi_2$  ( $C_\alpha C_\beta S_\gamma S_\gamma'$ ) and  $\chi_2'$  ( $S_\gamma S_\gamma' C_\beta' C_\alpha'$ ) dihedral angles around  $-90^\circ$  and  $\chi_3$  ( $C_\beta S_\gamma S_\gamma' C_\beta'$ ) close to  $+90^\circ$ .<sup>18,23,24,26,27</sup> This is affirmed in the present study, where over 96% of aCSDns were found to be right-handed staples (see Fig. 6A, B, E and F).

A naive understanding of disulfide torsion would suggest that the staple conformation, with staggered  $\chi_2$  and  $\chi_2'$  dihedral angles, should be a minimum on the torsional potential energy surface (PES). However, as seen in Fig. 5B, the potential energy well corresponding to the staple conformation ( $-60^\circ$ ,  $-60^\circ$ ) is much shallower than the other local minima on the calculated PES and is better described as a plateau on the surface. In fact, it is not seen at all in our calculated surface for  $\chi_3$  below  $100^\circ$ , the mean  $\chi_3$  value for aCSDns.<sup>22</sup>

This unexpectedly high torsional energy of the staple region of the PES is due to additional sources of structural strain affecting disulfides adopting the staple conformation. In NHB sites of canonical antiparallel  $\beta$ -sheet, the  $C_\alpha$  atoms of the two cross-strand residues are separated by roughly 4.5 Å. A disulfide conformation needs to have a comparable  $C_\alpha-C_\alpha$  separation in order to bridge between the two strands. Fig. 5C shows a contour plot for the  $C_\alpha-C_\alpha$  distance as the disulfide is twisted around  $\chi_2$  and  $\chi_2'$  (with  $\chi_3$  held at  $100^\circ$ ). A  $C_\alpha-C_\alpha$  separation of 4.5 Å corresponds to the cyan and green regions of this plot. However, as indicated by the red outline in Fig. 5C, aCSDns actually adopt conformations with lower  $C_\alpha-C_\alpha$  distances than those found in canonical  $\beta$ -sheet. In fact, the average  $C_\alpha$  separation of the aCSDns in our database is  $4.1 \pm 0.2$  Å. Comparison of Fig. 5B and C reveals that that this compression of the  $C_\alpha-C_\alpha$  separation gives conformations closer to where the staple minimum would be expected to sit on the PES. However, the

would be expected to be particularly vulnerable to reduction. (C) Contour plot of the variation in  $C_\alpha-C_\alpha$  distance (in Å) with torsion around  $\chi_2$  and  $\chi_2'$  ( $\chi_3 = 100^\circ$ ). The optimal  $C_\alpha-C_\alpha$  distance for the NHB site in antiparallel  $\beta$ -sheet is 4.5 Å (on the cyan-green contour), however most aCSDns adopt the slightly shorter  $C_\alpha-C_\alpha$  distances associated with the light and light-blue regions of the plot. (D) Contour plot of the variation in the dihedral angle between the  $C_\alpha-C_\beta$  bonds of the two Cys residues (in degrees) with torsion around  $\chi_2$  and  $\chi_2'$  ( $\chi_3 = 100^\circ$ ). For residue pairs in regular  $\beta$ -sheet, this dihedral angle should ideally be below  $60^\circ$  (light and dark blue regions of the plot). In (B–D) the regions of conformational space occupied by right-handed staple aCSDns are marked in red and those occupied by left-handed cis aCSDns outlined with a dashed magenta line. Note: in (B–D) axis labels are shown for right-handed conformations. For left-handed conformations the axes must be reversed, as in Fig. 6A.





**Fig. 6** Disulfide conformations adopted by aCSDn motifs showing the homogeneity of right-handed aCSDn conformations and the heterogeneity of left-handed structures. (A) and (B) show the  $\chi_2$  and  $\chi_2'$  dihedral angles adopted by left- and right-handed aCSDns, respectively. Conformations of right-handed disulfides with  $85^\circ \leq \chi_3 \leq 115^\circ$  (representing >95% of right-handed aCSDns) are superimposed on a contour plot of the potential energy surface (PES) for torsion around  $\chi_2$  and  $\chi_2'$  ( $\chi_3 = 100^\circ$ ). In most cases these disulfides are associated with medium-energy (turquoise) regions of the surface. Due to the much wider range of  $\chi_3$  dihedral angles seen for left-handed conformations, it is not instructive to superimpose these conformations on a PES plot for a single value of  $\chi_3$ , see Methods. In (C) and (D), representative structures of the various disulfide conformations and sub-motifs adopted by aCSDns are superimposed. (C) Disulfides with left-handed conformations contrasted with a right-handed staple shown in blue (PDB: 1rgx Cys 35 to Cys 88) (a single representative backbone has been used to improve clarity): left-handed staple (green, PDB: 2bba Cys 61 to Cys 184), left-handed cis from upper left of map (A) (red, PDB: 3flp Cys 37 to Cys 103), left-handed cis from lower right of map (A) (yellow, PDB: 2ny0 Cys 1130 to Cys 1159). (D) Disulfides with right-handed conformations: normal true-aCSDn (blue, PDB: 1rgx Cys 35 to Cys 88), true-aCSDn adjacent to an A-bent bulge (green, PDB: 1du5 Cys 51 to Cys 61), end-aCSDn (yellow, PDB: 2fkl Cys 133 to Cys 187),  $\beta$ -bridge-aCSDn (red, PDB: 1ivy, Cys 60 to Cys 334). All structures are seen to adopt very similar right-handed staple conformations. Parts (C) and (D) prepared using Molscript<sup>28</sup> and Raster3D.<sup>29</sup> (E) and (F) show relative populations left- and right-handed aCSDns in our dataset and their distribution amongst the various aCSDns sub-motifs and disulfide bond conformations.



short  $C_{\alpha}$ - $C_{\alpha}$  distance also brings the  $H_{\alpha}$  atoms of the Cys residues into close proximity, causing steric repulsion. This destabilises the staple conformation, reducing the depth of the minimum on the PES. Most staple disulfides are therefore in a medium-energy (cyan) region of the PES with torsional energies of 10 to 15  $\text{kJ mol}^{-1}$  above the minimum (see Fig. 6B). Hence, the staple conformation represents a trade-off between torsional strain of the disulfide bond on the one hand and steric repulsion between  $H_{\alpha}$  atoms plus strain in the  $\beta$ -sheet (due to the non-optimal  $C_{\alpha}$ - $C_{\alpha}$  distance) on the other.

However, comparison of Fig. 5B and C reveals that there are other regions of conformational space which also satisfy the conditions of relatively low torsional energy and a  $C_{\alpha}$ - $C_{\alpha}$  separation of  $\sim 4.5$  Å. In particular, conformations with one of  $\chi_2$  or  $\chi_2'$  approximately  $-60^\circ$  and the other close to  $+40^\circ$  (the hook conformation<sup>8</sup>) are significantly less strained than those in the staple region, and have  $C_{\alpha}$ - $C_{\alpha}$  distances much closer to 4.5 Å. However, there is an additional factor which must be satisfied for a conformation to be compatible with cross-strand residues in regular  $\beta$ -sheet: the  $C_{\alpha}$ - $C_{\beta}$  bonds of the two Cys must be roughly parallel. Fig. 5D shows the variation in the dihedral angle between the two  $C_{\alpha}$ - $C_{\beta}$  bonds as  $\chi_2$  and  $\chi_2'$  are varied ( $\chi_3$  is again held at  $100^\circ$ ). This shows that for hook conformations the dihedral angle between the  $C_{\alpha}$ - $C_{\beta}$  bonds is generally above  $90^\circ$ . Hence, the staple is also the only conformation which satisfies all three criteria. It is therefore not surprising that almost all true- and  $\beta$ -bridge-aCSDNs are staples.

With no hydrogen bonds on one side of the disulfide, it might be expected that end-aCSDNs would not be under such rigid conformational constraint. The conditions on the  $C_{\alpha}$ - $C_{\alpha}$  distance and the dihedral angle between the  $C_{\alpha}$ - $C_{\beta}$  bonds need not apply to end-aCSDNs. Nevertheless, almost all end-aCSDNs are also found in the staple region. As can be seen in Fig. 6B, the distribution of their conformations is slightly different to that of true- and  $\beta$ -bridge-aCSDNs. The end-aCSDn population forms an arc slightly closer to the centre of the plot (red triangles in Fig. 6B). This corresponds to a relaxation of the  $C_{\alpha}$ - $C_{\alpha}$  distance from (on average)  $3.97 \pm 0.14$  Å for true-aCSDNs to  $4.27 \pm 0.23$  Å for end-aCSDNs. Nevertheless, the conformations adopted by aCSDNs are very consistent. This can be seen in Fig. 6D, where representative structures of right-handed true-, end- and  $\beta$ -bridge-aCSDn sub-motifs have been superimposed along with an example of a true-aCSDn immediately adjacent to an A-bent bulge. The structures were chosen to give a range of points across the population in Fig. 6B, however the disulfides still superimpose closely upon one other. It is likely that slight twisting in the  $\chi_1$  and  $\chi_1'$  dihedral angles compensates for the variation in  $\chi_2$  and  $\chi_2'$  seen, particularly for end-aCSDNs.

155 aCSDn structures with left-handed conformations are also found in our dataset (Fig. 6A). Thus, although aCSDNs predominantly adopt the right-handed staple conformation, they do not do so exclusively, as previously thought.<sup>27</sup> Furthermore, not all of these adopt the staple conformation. While a few examples of left-handed end-aCSDNs are found in the staple region, for most left-handed true-aCSDNs one of  $\chi_2$  or  $\chi_2'$  is approximately  $110^\circ$ , and the other is close to  $0^\circ$  (ranging from  $-60^\circ$  to  $+60^\circ$ ). We refer to this as the left-handed cis

conformation. Fig. 5C and D reveal that this combination of dihedral angles (marked in magenta) produces a  $C_{\alpha}$ - $C_{\alpha}$  separation suitable for cross-strand  $\beta$ -partners, and can also give a dihedral angle between the  $C_{\alpha}$ - $C_{\beta}$  bonds below  $60^\circ$ . However, this conformation generally corresponds to high torsional energy regions of the potential energy surface (yellow and green in Fig. 5B). Indeed, wherever the torsional energy falls below 15  $\text{kJ mol}^{-1}$  in these regions, the angle between the  $C_{\alpha}$ - $C_{\beta}$  bonds rises above  $60^\circ$ . (Note: for many aCSDNs adopting the cis conformation,  $\chi_3$  is much lower than  $100^\circ$ . These disulfides have even higher torsional energies than suggested by Fig. 5B.<sup>22,44</sup>) This means that significant levels of strain must be accommodated for a disulfide to adopt the cis conformation, either as torsional strain in the disulfide bond or as twisting strain in the backbone of the  $\beta$ -ladder. Thus many left-handed aCSDNs are highly strained and are likely to be more reactive than their right-handed analogues.

The three conformations adopted by left-handed aCSDNs are superimposed in Fig. 6C, along with a right-handed staple aCSDn (blue) for comparison. The left-handed staple conformation (upper right quadrant of Fig. 6A) is shown in green, whereas the two rarer cis populations are depicted in red and yellow. The red conformation in Fig. 6C corresponds to the upper left group in Fig. 6A and the yellow to the lower right group. Although only 3% of the aCSDNs in our dataset are left-handed, left-handed disulfides are represented in 11% of the aCSDn clusters. The presence of left-handed aCSDNs in over 25 unrelated protein structures indicates they are unlikely to be simply the result of modelling errors. While 54% of the left-handed aCSDNs in our database are found in clusters where over 95% of the disulfides are left-handed, 21% are in clusters where the disulfides are split between left- and right-handed conformations (20–95% left-handed). In other words, there are some aCSDn clusters which are significantly enriched in left-handed disulfides; for example: in the ligand-binding domain of ephrin receptors, the disulfide connecting Cys 61 to Cys 184 adopts a left-handed staple conformation in 48 of 49 instances (e.g. PDB: 2bba). For other clusters, however, there is evidence to suggest that the disulfide may alternate between left- and right-handed conformations, with the higher energy left-handed forms potentially being more readily reduced. A particularly interesting example is the disulfide which stores the resolving Cys 213 in the 2-Cys archaeal peroxiredoxin.<sup>45</sup> Cys 213 normally forms a right-handed staple true-aCSDn with Cys 207 under oxidising conditions ( $\text{pH} \geq 7.5$ ; PDB: 2cv4, 3a2w, average DTE =  $11.8 \pm 0.6$   $\text{kJ mol}^{-1}$ ), however at lower pH (PDB: 3a5w) the disulfide is generally found in a much higher energy left-handed-cis conformation (average DTE =  $20 \pm 2$   $\text{kJ mol}^{-1}$ ); in many structures the strain of this conformation has forced the disulfide bond to break.<sup>46</sup> This could be either due to radiative damage during structure solution or the result of physico-chemical events.

Our database also contains a very small number of disulfides (13) adopting a right-handed version of the cis conformation. Almost all of these have  $\chi_3 < 85^\circ$  (often significantly less) and hence are not shown in Fig. 6B. Their conformations, along with all other right-handed aCSDNs with  $\chi_3$  outside the range of  $85^\circ$  to



115°, can be found in Fig. S3.† As these right-handed cis disulfides represent only 0.2% of aCSDns, their numbers are currently too small to assess their properties and biological relevance.

**(b) Dependence of disulfide torsional strain on embeddedness.** As in previous works,<sup>11,22,44</sup> we assessed the strain experienced by each aCSDn in our database due to torsion around the dihedral angles of the disulfide bond. In particular, we investigated the dependence of the torsional strain of the disulfide on the embeddedness of the motif. Using the PES, the average torsional strain for each aCSDn sub-motif was calculated. With respect to longitudinal embeddedness, the right-handed true-, end- and  $\beta$ -bridge-aCSDns have similar average disulfide torsional energies (DTE, strain in  $\chi_2$ ,  $\chi_3$  and  $\chi_2'$ ) of 12 kJ mol<sup>-1</sup> (see Table 1). Similarly, there is no variation in DTE with lateral embeddedness.

It is also informative, however, to look at the distribution of disulfide torsional strains for aCSDns with different levels of LonE. Although true- and end-aCSDns have the same mean disulfide energies, their energy distributions are quite different (see Fig. 7A). For the end-aCSDn population, the torsional energy distribution rises gradually to the modal value of 12.5–15 kJ mol<sup>-1</sup> then falls rapidly, while for the true- and  $\beta$ -bridge-aCSDn populations the distribution rises quickly to its maximum of 10–12.5 kJ mol<sup>-1</sup> then slowly tails away. Thus, despite the greater conformational freedom available to end-aCSDns, their structures tend to be more highly strained than those adopted by true- and  $\beta$ -bridge-aCSDns. Some end-aCSDn structures do utilise the additional conformational freedom available to them to reduce their disulfide energies, however, resulting in the lower mean disulfide torsional strain of the sub-motif. An estimate can also be made of the contributions of the strain in the  $\chi_1$  and  $\chi_1'$  dihedral angles to the total torsional strain (TTE). It is significant that, while the  $\chi_1$  and  $\chi_1'$  contributions to the TTE are comparatively small for true- and  $\beta$ -bridge-aCSDns (on average +1.8 kJ mol<sup>-1</sup>), they are more than twice as large for end-aCSDns (average +3.7 kJ mol<sup>-1</sup>). This can be clearly seen in the total energy histograms in Fig. 7B, where the populations of true- and  $\beta$ -bridge-aCSDns have a reasonably tight energy distribution ( $\sigma = 2.3$  and 1.9 kJ mol<sup>-1</sup>, respectively). Very few true- and  $\beta$ -bridge-aCSDns have total energies higher than 20 kJ mol<sup>-1</sup>. In contrast the end-aCSDn population has a broader distribution ( $\sigma = 4$  kJ mol<sup>-1</sup>) with a significant

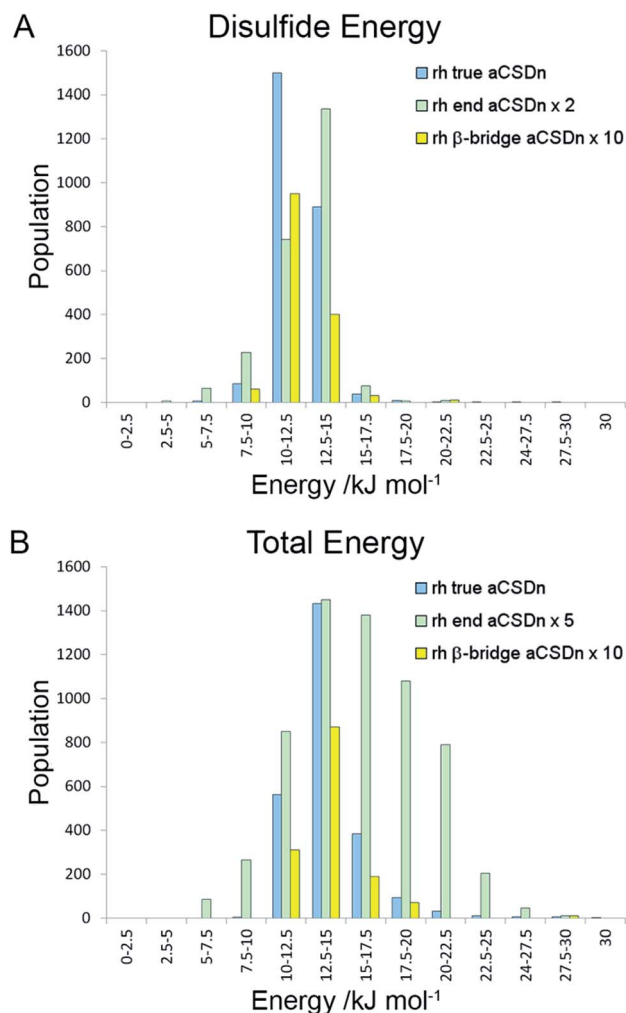


Fig. 7 Distribution of torsional energies for right-handed disulfides in aCSDn motifs. (A) The disulfide torsional energy (DTE) distribution *i.e.* the strain in  $\chi_2$ ,  $\chi_3$  and  $\chi_2'$ . (B) The total torsional energy (TTE) distribution *i.e.* the strain in  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_2'$  and  $\chi_1'$ . The populations for end- and  $\beta$ -bridge aCSDns have been scaled, as described in the legend, to allow easy comparison on the same set of axes.

proportion of disulfides with total energies in the 20–22.5 kJ mol<sup>-1</sup> bin. This supports the suggestion that increased torsion in  $\chi_1$  and  $\chi_1'$  compensates for the broader range of  $\chi_2$  and  $\chi_2'$

**Table 1** Distribution of the torsional strain between the disulfide dihedral angles for left- and right-handed aCSDn motifs and conformations. Torsional strain in the central  $\chi_2$ ,  $\chi_3$  and  $\chi_2'$  dihedral angles is listed as DTE and the overall torsional strain is denoted TTE. Average energies and standard deviations are given in kJ mol<sup>-1</sup>

		Left handed						Right handed							
		No.	DTE		$\chi_1\chi_1'$		TTE		No.	DTE		$\chi_1\chi_1'$		TTE	
LonE	Conf.		Ave	$\sigma$	Ave	$\sigma$	Ave	$\sigma$		Ave	$\sigma$	Ave	$\sigma$	Ave	$\sigma$
True	Staple	38	13.8	2.8	10	10	24	11	3138	12.2	1.4	1.8	2.0	14.0	2.3
	cis	51	24	7	11	4	34	7	5	60	40	13	8	70	40
End	Staple	50	12.4	0.7	4.4	2.7	17	3	1406	12.4	2.3	3.7	2.9	16	4
	cis	14	20	6	8	10	28	13	8	20	11	3.0	1.3	23	10
$\beta$ -Bridge	Staple	1	14.1		0.1		14.2		150	12.1	1.5	1.9	1.6	13.9	1.9
	cis	1	21.4		20.6		42.0		0						



values seen in the end-aCSDn population (Fig. 5B), resulting in a correspondingly higher contribution of  $\chi_1$  and  $\chi_1'$  to the TTE. Thus, although end-aCSDns were expected to be less strained than the corresponding true sub-motifs, in fact they have (on average) higher total strain energies.

As expected, the strain of incorporating a left-handed aCSDn into a  $\beta$ -sheet is much higher than for the corresponding right-handed motifs. Left-handed true-aCSDns have high levels of both DTE and ( $\chi_1$ ,  $\chi_1'$ ) strain when either the staple or cis conformations are adopted (see Table 1). The average TTEs in both cases were over  $25 \text{ kJ mol}^{-1}$ , indicating disulfides which are likely easily reduced. Some relaxation of the disulfide conformation was seen for left-handed end-aCSDns resulting in lower total energies. For disulfides which adopt the cis conformation this relaxation was only seen in the  $\chi_1$  and  $\chi_1'$  dihedral angles, giving total energies which were still  $\sim 25 \text{ kJ mol}^{-1}$ . In contrast, in left-handed staple end-aCSDns all five dihedral angles are relaxed, resulting in conformations with a similar level of strain to the right-handed end-aCSDns (DTE  $12.6 \text{ kJ mol}^{-1}$ ; TTE  $17.0 \text{ kJ mol}^{-1}$ ). Very high average DTE and TTEs were also seen for the small number of aCSDns adopting the right-handed cis conformation. The extremely high average values seen for right-handed cis true-aCSDns ( $60$  and  $70 \text{ kJ mol}^{-1}$  respectively) likely indicates that modelling errors are present in at least some of these structures.

**(c) Interplay of disulfide handedness and conformation with  $\beta$ -sheet twist and shear.** While it is understandable that aCSDns with left-handed cis conformations are rare, it is not immediately clear why the left-handed staple should be highly strained and hence disfavoured. In order to investigate the principles underlying this phenomenon, *ab initio* quantum chemical calculations were performed on model compounds of both left- and right-handed aCSDns (Fig. 8). The full geometry optimisations yielded structures with mirror image disulfide conformations (right-handed:  $\chi_2 = -93^\circ$ ,  $\chi_3 = 98^\circ$ ,  $\chi_2' = -93^\circ$ ; left-handed:  $\chi_2 = 94^\circ$ ,  $\chi_3 = -97^\circ$ ,  $\chi_2' = 96^\circ$ ). The  $\chi_1$  and  $\chi_1'$  dihedral angles were also very close to their optimal values (the values required for a staggered arrangement around the  $C_\alpha$ - $C_\beta$  bond) of  $-60^\circ$  for right-handed staples and  $180^\circ$  for left-handed staples ( $-54^\circ$  and  $193^\circ$  respectively). This reveals that the preference for the right-handed conformation is not an inherent property of the disulfide-bonded cysteine residues themselves. In other words, left-handed aCSDns are not disfavoured due to a conflict between the handedness of the disulfide and the chirality of the L-amino acids.

However, when the relative twisting and shearing of the two Cys residues in the model compound calculations are compared, the reason for the preference in handedness becomes clear. The optimised geometries (shown in Fig. 8) reveal that introduction of a right-handed staple disulfide (on the right in the figure) leads to a right-handed twist of  $+24.6^\circ$  in the  $\beta$ -ladder (see also Table 2). In contrast, a left-handed staple disulfide produces a left-handed twist of  $-32.3^\circ$ . In a data mining analysis of 106 NHB sites in antiparallel  $\beta$ -ribbons (which our calculations model), Ho and Curmi<sup>47</sup> found that the preferred backbone dihedral angles ( $\phi$  and  $\psi$ ) are associated with a twist of  $+25 \pm 16^\circ$  between the residues. Thus a right-

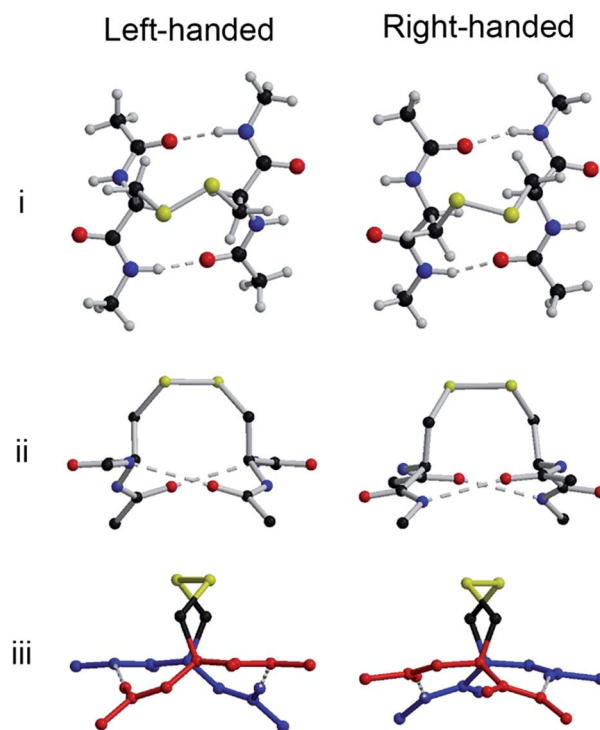


Fig. 8 Optimised geometries for the aCSDn model compound. Compounds with a left-handed staple disulfide induce a left-handed twist in the sheet, whereas those with a right-handed staple disulfide induce a right-handed sheet twist. (i) View looking down onto the face of the sheet. (ii) View from the side looking along the strands. Atom sizes have been reduced and hydrogen atoms removed to improve clarity. (iii) View from the side across the strands. The strands have been given different colours to visualize the sheet twist. Figure prepared using Molscript<sup>28</sup> and Raster3D.<sup>29</sup>

handed staple aCSDn can be inserted into an antiparallel  $\beta$ -sheet without causing any additional twisting strain. On the other hand, a left-handed staple disulfide cannot be inserted

Table 2 Twisting and shearing of residues across NHB sites in antiparallel  $\beta$ -sheet. Results for "normal NHB sites" (*i.e.* sites with no disulfide present) are taken from Ho and Curmi.<sup>47</sup> These authors only presented data for NHB sites in mid-sheet and  $\beta$ -ribbon LatE contexts. The data for  $\beta$ -ribbon contexts have been presented here, as these give the best comparison with both the model compounds and the majority of PDB structures

	Twist	Shear
Normal NHB sites	$+25 \pm 16^\circ$	$+0.80 \pm 0.25 \text{ \AA}$
<b>Right-handed aCSDn</b>		
Model compound	$+24.6^\circ$	$+1.04 \text{ \AA}$
Average from PDB	$+25 \pm 8^\circ$	$+0.9 \pm 0.3 \text{ \AA}$
<b>Left-handed aCSDn</b>		
Model compound	$-32.3^\circ$	$+0.56 \text{ \AA}$
Average from PDB		
Negative twist	$-25 \pm 14^\circ$	$+0.77 \pm 0.26 \text{ \AA}$
Positive twist	$+16 \pm 9^\circ$	$+0.72 \pm 0.27 \text{ \AA}$



into an antiparallel  $\beta$ -sheet without significant distortion of either the sheet or the disulfide.

The optimised model compounds also reveal differences in the way right- and left-handed disulfides affect the shear of  $\beta$ -sheets. The data mining study of Ho and Curmi showed that residues in NHB sites of antiparallel  $\beta$ -sheet exhibit a shear of  $+0.80 \pm 0.25$  Å towards the C-terminus of the strand. This arrangement allows weak hydrogen bonding interactions between the  $H_{\alpha}$  atoms and the carbonyl oxygens (see Fig. S1C†). Our *ab initio* calculations reveal that inclusion of a right-handed staple aCSDn increases this shear to  $+1.04$  Å, while a left-handed conformation reduces it to  $+0.56$  Å. Thus the two disulfide chiralities have roughly equal and opposite effects on the sheet shear.

To assess the effect of aCSDns on the backbone conformations of protein structures, we calculated the relative twist and shear between the two Cys residues of each aCSDn structure in our database. These were compared with the data mining results of Ho and Curmi<sup>47</sup> for residue pairs in NHB sites with no connecting disulfide (hereafter referred to as “normal” residue pairs or “normal” NHB sites for simplicity). The average twist for the right-handed staple aCSDns in our database is  $+25^\circ$  with a standard deviation of  $8^\circ$  ( $n = 4694$ ). This is comparable to the average twist of  $+25^\circ \pm 16^\circ$  for normal NHB sites in  $\beta$ -ribbon structures (which give the best comparison with most structures in our database).<sup>47</sup> It appears, however, that the presence of the disulfide somewhat restricts the allowed range of the twist, as reflected in the standard deviation. The average sheet shear of right-handed staple aCSDns is  $+0.9 \pm 0.3$  Å, intermediate between the calculated value of  $+1.04$  Å for the model disulfide and the value from the data mining study for normal residues pairs of  $+0.80 \pm 0.25$  Å, however all three results are consistent within the error margins. Ho and Curmi also reported sheet twist and shear for NHB sites in the interior of  $\beta$ -sheets. This context corresponds to true-aCSDns (subclass 1) with mid-sheet LatE. Again the twist we observed for this subset of aCSDns ( $+17 \pm 8^\circ$ ) agreed well with Ho and Curmi's results for normal NHB sites ( $+15 \pm 10^\circ$ ), while the shear of  $+0.95 \pm 0.22$  Å was intermediate between our modelled value and that obtained in Ho and Curmi's data mining study ( $+0.85 \pm 0.25$  Å). In general, the relative twist of the cysteine residue pairs in our dataset was not found to depend on either the LonE or LatE context of the disulfide. One exception was the cluster of TLSP structures which dominate the population of end-aCSDn-2s in our dataset. In these structures the  $\beta$ -strands of the Cys pairs are more highly twisted ( $+34 \pm 5^\circ$ ) than other aCSDn structures. The Cys pairs of these TLSP disulfides are also more highly sheared than those of most other aCSDns,  $+1.06 \pm 0.25$  Å.

The relative shear of the Cys residues also does not appear to have any clear correlation with the level of LonE, however there may be some correlation with the LatE if the TLSP structures are excluded. Disulfides on edge strands seem to be associated with slightly higher levels of shear ( $+0.93 \pm 0.30$  Å) than those in the middle of  $\beta$ -sheets ( $+0.84 \pm 0.28$  Å), or  $\beta$ -ribbon structures ( $+0.82 \pm 0.30$  Å). The observed shears for  $\beta$ -ribbon and mid-sheet disulfides in our database are also consistent with the findings of Ho and Curmi ( $+0.80 \pm 0.25$  Å and  $+0.85 \pm 0.25$  Å respectively).

These authors did not look at structures on the edges of  $\beta$ -sheets. We conclude, therefore, that the presence of a right-handed staple aCSDn does not result in any significant perturbation of the “normal” twist and shear of the surrounding  $\beta$ -sheet context.

Whereas right-handed aCSDns fit well with the natural right-handed twist of  $\beta$ -sheets, our quantum chemical calculations reveal that left-handed aCSDns prefer to be associated with a significant left-handed twist. They can therefore only be accommodated by either twisting the sheet away from its natural right-handed tendency or by straining the disulfide. Both modes of accommodation are observed in our dataset. Approximately half of the left-handed aCSDns are associated with left-handed sheet twists. These mostly adopt the staple conformation. The average twist of these structures is  $-25 \pm 14^\circ$ , and their shear is  $+0.77 \pm 0.26$  Å. This shear is slightly lower than that of right-handed aCSDns, but in the same direction. The remaining left-handed aCSDns are found in sheets with right-handed twists. They are predominantly true-aCSDns, however all levels of LonE (and LatE) are represented. In order to accommodate the existing right-handed twist of the sheet, the left-handed disulfide is placed under significant torsional strain. This strain can be accommodated by either adopting unfavourable  $\chi_1$  and  $\chi_1'$  dihedral angles for a left-handed staple disulfide; or by twisting into the higher energy left-handed cis conformation. These structures are found to have a wide range of right-handed sheet twists although the magnitude is generally smaller (average:  $+16 \pm 9^\circ$ ) than those of right-handed aCSDns. As predicted by calculations on the model compound, the shear between Cys residue pairs in these left-handed disulfide structures ( $+0.72 \pm 0.27$  Å) is lower than when the disulfide is not present ( $+0.80 \pm 0.25$  Å).

### C Evidence of aCSDn redox activity

We searched for biochemical and physicochemical evidence of redox activity for the aCSDns identified in this study. For a disulfide to be confirmed as redox-active based on evidence from this extensive literature search, we used the fairly strict criteria that: (1) redox activity has been experimentally demonstrated specifically for the aCSDn of interest; and (2) that the disulfide has been shown to perform a functional role under physiological conditions. We also examined crystallographic data for evidence of aCSDn redox activity. This included: examples of both oxidised and reduced structures of the aCSDn in our dataset;  $S_{\gamma}$ - $S_{\gamma}$  distances between 2.2 and 3.6 Å, indicative of a mixed oxidation state of the disulfide; and non-unity occupancy of the atoms of the Cys residues.

We found experimental confirmation of redox activity at the level of the specific disulfide for 5% of the 254 aCSDn clusters identified in this work. Details of the properties of these disulfides, along with literature references describing their redox activity, are shown in Table 3; further information on the biological functions of these disulfides can be found in Table 4. Further aCSDn clusters for which evidence of disulfide reduction is found in X-ray crystal structures are also included in Table 3. In total, 13% of the aCSDn clusters have experimental evidence supporting redox activity. Importantly, the confirmed





**Table 3** Disulfide properties of the aCSDn clusters mentioned in this paper. In each case a representative disulfide is given along with the number of oxidised structures and their average total torsional energy, the LonE and LatE levels observed and the disulfide conformation(s) adopted. Also shown is an indication of whether the disulfide is required for the structure of the fold, a description of the taxa in which the disulfide cluster is found and any evidence for redox activity (either the reference of a publication in which it is reported or a PDB code in which the disulfide is reduced or in a mixed oxidation state). A complete list of disulfide properties and protein fold data for all aCSDn clusters can be found in Table S1

Cluster name <sup>a</sup>	Representative disulfide	No. Ox.	TTE (kJ mol <sup>-1</sup> )		LonE <sup>b</sup>	LatE <sup>b</sup>	Conf. <sup>c</sup>	Non Str. <sup>d</sup>	Taxonomic conservation <sup>e</sup>	Proof of redox activity <sup>f,g</sup>
			Ave	$\sigma$						
DsbD, N-terminal domain	1l6p A 103–109	2	14.3	0.8	True	1	rst	Yes	Bacteria	12
Glutathione peroxidase	3dww A 47–95	2	10.4	0.6	End	0	rst	Yes	Euglenozoa	37
Archaeal peroxiredoxin	2cv4 A 207–213	25	19	15	True	0	rst/lcs	Yes	Archaea	45
Bacterioferritin comigratory protein	2cx4 G 49–54	6	14.1	1.6	True/end	0	rst	Na	Bacteria + archaea	48
Cytokine receptors #1 <sup>g</sup>	2c4f U 186–209	50	14.4	2.4	True	0/1	rst	Yes	Euteleostomi + viruses	14
FKBP13	1u79 A 106–111	3	16.7	2.3	True (A-bent)	0	rst	Yes	Tracheophyta	34
Nucleoside triphosphate diphosphohydrolase #1	3agr A 234–244	2	16.2	0.6	True	0	rst	Yes	Alveolata	49
Thioredoxin (non-catalytic Cys)	2ifq A73–B73	2	14.2	0.3	$\beta$ -Bridge	0	rst	Na	Metazoa	50
Integrin $\alpha_2$ calf-2 domain	3ije A 822–884	3	14.7	1.2	True/end	1	rst	Yes	Metazoa	38
Gelsolin/CapG/adseverin	1kcc A 188–201	6	14.6	1.0	True	1 <sup>h</sup>	rst	Yes	Euteleostomi	51
Ephrin receptors, ligand binding domain <sup>i</sup>	2bba A 61–184	49	17.5	4.4	End	0	lst	Yes	Metazoa	52
Galectin CG-1B	3dnt A7–B7	1	11.0		True	2	rst	Yes	<i>Gallus gallus</i>	53 and 54
Translationaly-controlled tumor protein	3ebm B172–C172	1	13.7		True	0 <sup>h</sup>	rst	Na	Chordata + plants	55
Collagen IV, NC1 domain	1t60 A 64–70	96	13.4	1.1	True	2	rst	Na	Metazoa	56
InaD with NorpA	1ihj A31–D6	2	12.9	1.1	True	1	rst	Na	<i>Drosophila</i>	40
<i>Clostridium</i> neurotoxins	1epw A 436–445	9	13.4	1.2	True/end	1	rst	Na	<i>Clostridium</i>	35 and 36
Envelope glycoprotein gp120 #1	2nxy A 126–196	17	19	4	True/end	0	rst	No	Primate lentiviruses	57
Envelope glycoprotein gp120 #2	2nxy A 218–247	31	15.6	2.0	End/ $\beta$ -bridge	0/1	rst	No	Primate lentiviruses	13
Envelope glycoprotein gp120 #4	2nxy A 385–418	32	22	11	True	1/2 <sup>h</sup>	rst/lcs	No	Primate lentiviruses	58
CD4, 1 <sup>st</sup> C2-set domain	1cdy A 130–159	23	18	11	True	0/1/2	rst/lcs	Yes	Primates + nuroids	59
Yellow fluorescent protein <sup>i</sup>	1h6r A 149–202	4	12.6	0.9	True	1	rst	Yes	Engineered	13
Green fluorescent protein <sup>i</sup>	1jcl A 147–204	4	12.9	1.2	End	1	rst	Yes	Engineered	58
Hydrophobins #1	1r2m A 14–26	13	12.9	1.2	True (A-bent)	1/2	rst	No	Dikarya	59
Thaumatin-like proteins <sup>k</sup>	2pe7 A 56–66	15	16	10	True (A-bent)	0	rst/lst	Na	Bacteria + eukaryota	1rqw A 56–66
Eukaryotic trypsin-like serine proteases <sup>l</sup>	1s5s A 136–201	936	17	4	True/end	1/2 <sup>h</sup>	rst	Yes	Vertebrates + some other metazoa	1y3w A 136–203
Bowman-Birk inhibitors	2g81 I 24–32	30	13.7	1.7	True/end	0/1/2	rst	Yes	Eudicoyl/dons, not liliopsida	2r33 A 24–32
$\beta$ -Lactamase-inhibitor protein-1, BLIP-I #2	3gmu B 109–131	29	14.5	1.8	True/end	1	rst	Yes	<i>Streptomyces</i>	
Glycosyl hydrolase family 7 #2	1ojk A 215–234	55	14	4	$\beta$ -Bridge	0	rst	Na	Fungi, amoebozoa, parabasalia	
Glycosyl hydrolase family 12, catalytic domain #1	2nlr A 5–31	42	15	3	True/end	1	rst/lcs	Na	Bacteria, archaea, fungi, plants, stramenopiles	
Peptidase A4 #1	2ifr A 141–148	4	14	3	True/end	0	rst	Yes	Pezizomycotina (not Eurotiomycetes)	
Serine carboxypeptidase-like proteins <sup>m</sup> #1	1ivy A 60–334	12	12.8	1.1	$\beta$ -Bridge	0	rst	Na	Eukaryota	
Amine dehydrogenases <sup>n</sup> , heavy chain	2bbk H 168–183	84	15.0	1.7	True/end/ $\beta$ -bridge	1	rst	Yes	<i>Proteobacteria</i> + <i>Actinobacteria</i>	1mda J 167–183
TNF ligand superfamily member 4, OX40L	2hew F 98–183	4	12.4	0.2	$\beta$ -Bridge	0	rst	Yes	Mammalia	
Pilus chaperones (PapD N-terminal domain-like) <sup>p</sup>	1z9s A 98–137	14	12.8	0.9	True	1/2	rst	Yes	Bacteria	1p5u A 98–137 <sup>p</sup>



Table 3 (Contd.)

Cluster name <sup>a</sup>	Representative disulfide	No. Ox.	TTE (kJ mol <sup>-1</sup> )		LonE <sup>b</sup>	LatE <sup>b</sup>	Conf. <sup>c</sup>	Non Str? <sup>d</sup>	Taxonomic conservation <sup>e</sup>	Proof of redox activity <sup>f</sup>
			Ave	$\sigma$						
Kunitz (STI) inhibitors	2qn4 B 140–144	15	14.2	2.3	True/end	0/1	rst/lst	Yes	Viridiplantae	60 <sup>g</sup>
$\beta$ -Lactoglobulin	1beb A 106–119	38	14.2	2.9	True	2	rst	Yes	Mammalia	61 <sup>h</sup>

<sup>a</sup> Where more than one unique aCSDn is found in a particular protein or protein type, these are distinguished by #1, #2, etc. according to their position in the protein sequence. <sup>b</sup> Levels of embeddedness shown in grey occur in less than 10% of structures. <sup>c</sup> Prefixes “r” and “l” indicate right- and left-handed conformations respectively. St = staple, cs = cis. Only conformations seen in at least 10% of structures are listed. <sup>d</sup> Can the protein fold exist without the disulfide? Yes = fold has been seen to exist without the disulfide; No = the disulfide is not in the central fold of the protein; No = no crystal structures of this fold without the disulfide have yet been reported. See Table S1 for more details. <sup>e</sup> A brief indication of the taxa in which the disulfide is conserved. Generally, the disulfide will be found in multiple (but not necessarily all) sub-taxa for the given taxon. A more in-depth analysis, including discussions of orthologs where the disulfide is not conserved, will be published elsewhere. <sup>f</sup> Numbers indicate literature references which describe the redox activity of the disulfide. PDB codes indicate structures in which the disulfide is reduced or in a mixed oxidation state. <sup>g</sup> Tissue factor; all type 2 cytokine receptors (interferon receptors (including in viruses), interleukin-10, -20, -22a2 receptors); some type 1 cytokine receptors (growth hormone receptor, interleukin-7 $\alpha$  and -13 $\alpha$  receptors). <sup>h</sup> Access to this disulfide is impeded along both edges. <sup>i</sup> A2, A4, B2 and B4. <sup>j</sup> Engineered. <sup>k</sup> Thaumatin, osmotin, zeamatin, PR5. <sup>l</sup> Trypsin, chymotrypsin, all kallikreins, azurocidin, acrosin, cathepsin G, myeloblastin, hepsin, plasminogen, plasminogen activators, all granzymes, venom serine protease, elastase, urokinase, trypsin, enteropeptidase, procarboxypeptidase A, duodenase, chymase 1 and 2, hepatocyte growth factor, complement factor D, coagulation factor XI. Disulfide is conserved in all vertebrata but only in some urochordata, arthropoda, mollusca and annelida. <sup>m</sup> Not vitellogenic carboxypeptidase-like protein or retinoid-inducible serine carboxypeptidase. <sup>n</sup> Methylamine dehydrogenase and aromatic amine dehydrogenase. <sup>o</sup> aafD, afab, agg(3)D, cafIm, cs3-1, essC, myfB, nfaE, psab, safB, sefB. <sup>p</sup> The disulfide in this structure is in a mixed oxidation state. <sup>q</sup> This disulfide has been shown to be a thioredoxin substrate, however the biological relevance, if any, of this reactivity has not yet been elucidated.

examples were distributed throughout the clusters identified, suggesting that redox behaviour is not restricted to a small group of homologous proteins but is a property of the motif, as indicated by previous computational results.<sup>22,44</sup> Significantly, whenever the redox activity of an aCSDn was specifically investigated, the disulfide was always demonstrated to be redox-active.

Although the numbers of confirmed redox-active aCSDns are small, the distribution of clusters with known redox activity amongst the aCSDn sub-motifs appears to reflect the overall cluster distribution shown in Fig. 3B. Hence, redox activity is not enriched or depleted for any of the sub-motifs. Confirmation of redox activity was found in 13 clusters which adopt only true-aCSDn sub-motifs, two clusters with only end-aCSDns, four clusters where the disulfides are split between “true” and “end” sub-motifs and one cluster where the disulfide is always a  $\beta$ -bridge-aCSDn (Table 3). The redox-active aCSDns identified in our database are found in structures with all three levels of LatE. The clusters with only end-aCSDn structures are found either on  $\beta$ -ribbons (glutathione peroxidase, see Table 3 for PDB codes) or edge strands (engineered green fluorescent protein). Of the true-aCSDn clusters with known redox activity, four are found only in  $\beta$ -ribbon contexts (FKBP13, archaeal peroxiredoxin, nucleoside triphosphate diphosphohydrolase #1 and translationally-controlled tumor protein), four are exclusively found on edge strands (DsbD, gelsolin, engineered yellow fluorescent protein and InaD with NorpA) and for two clusters all structures are in the middle of  $\beta$ -sheets (chicken galectin CG-1B and the NC1 domain of collagen IV). Additionally, one cluster is split between  $\beta$ -ribbon and edge strand contexts (cytokine receptors #1), another is split between edge-of-sheet and mid-sheet contexts (HIV and SIV envelope glycoprotein gp120 #4) and for a final cluster all three levels of LatE are seen (CD4, 1<sup>st</sup> C2-set domain). These heterogeneous clusters may reflect conformational lability. Two of the redox-active clusters with examples of both true- and end-aCSDn sub-motifs are found exclusively in  $\beta$ -ribbons (bacterioferritin comigratory protein and HIV/SIV envelope glycoprotein gp120 #1), and two are only on edge strands (*Clostridium* neurotoxins and integrin  $\alpha$ , calf-2 domain). The  $\beta$ -bridge-aCSDn structure (thioredoxin non-catalytic Cys) is in a  $\beta$ -ribbon LatE context. Again, the distribution of redox-active clusters amongst the different LatE contexts appears to roughly reflect the overall distribution within the dataset.

The confirmed redox-active aCSDns in our dataset include both intramolecular and intermolecular disulfides. Of the four intermolecular aCSDns, three involve identical residues of two monomers reacting to produce a homodimer (thioredoxin non-catalytic Cys, chicken galectin CG-1B and translationally-controlled tumor protein). Redox-controlled dimerization regulates the activation or deactivation of these proteins. For example, dimerization of translationally-controlled tumor protein *via* the oxidation of a pair of Cys 172 residues (*e.g.* PDB: 3ebm) is critical for its cytokine-like activity,<sup>54</sup> stimulating allergic inflammation, and is likely to also be important for the growth hormone activity of this protein.<sup>62</sup> Our dataset also contains one example of a redox-active disulfide-mediated

Table 4 Biological functions of known redox-active aCSDns

Cluster	Function	Biological role
DsbD, N-terminal domain	Catalytic	Redox homeostasis
Glutathione peroxidase	Catalytic	Redox homeostasis
Archaeal peroxiredoxin	Catalytic	Redox homeostasis
Bacterioferritin comigratory protein	Catalytic	Redox homeostasis
Cytokine receptors #1	Switching	Switching between signalling and coagulation functions
FKBP13	Switching	Activation/inactivation of isomerase functionality
Nucleoside triphosphate diphosphohydrolase #1	Switching	Activation/inactivation of NTPDase functionality
Thioredoxin (non-catalytic Cys)	Switching	Activation/inactivation of thioredoxin functionality
Integrin $\alpha$ , calf-2 domain	Switching	Activation/inactivation of integrin functionality
Gelsolin/CapG/adseverin	Switching	Allows different functions in different cellular compartments
Galectin CG-1B	Switching	Inflammatory and immune response
Translationally-controlled tumor protein	Switching	Immune response
Collagen IV, NC1 domain	Switching	Postulated to be immune response
InaD with NorpA	Switching	Proposed to help maintain the signalling complex under oxidising conditions
<i>Clostridium</i> neurotoxins	Switching	Cell entry
Envelope glycoprotein gp120 #1	Switching	Cell entry
Envelope glycoprotein gp120 #4	Switching	Cell entry
CD4, 1 <sup>st</sup> C2-set domain	Switching	Cell entry
Yellow fluorescent protein	Switching	Redox sensing
Green fluorescent protein	Switching	Redox sensing

heterodimer. Kimple, *et al.* showed that the reaction of Cys 31 in the N-terminal PDZ domain of the scaffolding protein “inactivation-no-after-potential D” (InaD) with a cysteine residue in the C-terminus of “no-receptor-potential A” (NorpA) forms a high-affinity complex which is essential for coordination of phototransduction in *Drosophila* eyes.<sup>56</sup>

To determine if these disulfides with experimental evidence for redox activity are representative of aCSDns in general, we compared the average torsional energies of disulfide clusters which have been shown to be involved in redox processes with those which have not yet been ascribed redox activity. (See Table 3.) If redox activity is an unusual function, it might be expected that torsional energies for known redox-active disulfides would be higher than for other aCSDns, however for only one experimentally confirmed redox-active cluster, the regulatory disulfide in FKBP13, is the strain more than one standard deviation above the average for the relevant sub-motif and disulfide conformation. In this “subclass 3” true-aCSDn (see Fig. 2) the average DTE and TTE are 13.3 and 16.7 kJ mol<sup>-1</sup> respectively. Another example of a high-energy redox-active disulfide is the end-aCSDn in HIV and SIV envelope glycoprotein gp120 (Cys 126 to Cys 196) which has an average TTE of 19 ± 4 kJ mol<sup>-1</sup>. (Note the higher average energy and standard deviation for end-aCSDns vs. true-aCSDns, see Table 1.) In contrast, the catalytic disulfide in glutathione peroxidase is at the lower end of the energy spectrum for aCSDn motifs, with a mean disulfide energy of 9.9 kJ mol<sup>-1</sup> and total energy of only 10.4 kJ mol<sup>-1</sup>. However, there is evidence in some crystal structures that aCSDns which are, on average, less strained may twist into higher energy conformations as a part of redox processes, for example the true-aCSDn of archaeal peroxiredoxins described earlier (PDB: 3a5w). A similar phenomenon is seen for the disulfide connecting Cys 120 and Cys 159 in the 1<sup>st</sup> C2-set

domain of CD4 (PDB: 1cdy) and for a second aCSDn in HIV and SIV envelope glycoprotein gp120 which connects Cys 385 to Cys 418 (PDB: 2nxy). As seen in Table 3, however, in most crystal structures of known redox-active aCSDns, the energies are close to the overall aCSDn average. This may be an indication that all aCSDns have the potential to be redox-active.

The functions and biological roles of the known redox-active aCSDns are shown in Table 4. It is interesting to note that many of these disulfides are involved either in maintenance of the redox potential of the relevant cellular compartment (redox homeostasis) or in switching responses to changes in this redox environment (immune responses, redox sensing, signalling, *etc.*). It appears that the intermediate levels of strain in aCSDns leave them uniquely poised to respond to fluctuations in the environmental redox potential.

Importantly, several of the aCSDns in our dataset are known to be thioredoxin (Trx) substrates. The Trx fold is a ubiquitous oxidoreductase fold. Although it is commonly believed that Trx acts as a general disulfide reductase, substrate specificity has been demonstrated both at the level of the protein target and individual disulfides. For example, high throughput screenings have identified barley  $\alpha$ -amylase/subtilisin inhibitor (BASI) as a Trx-h2 target in barley seeds.<sup>63</sup> BASI contains two disulfides but Trx-h2 preferentially reduces the aCSDn connecting Cys 144 to Cys 148 over the non-forbidden Cys 43 to Cys 90 disulfide (see Fig. 4, BASI is shown in magenta).<sup>60</sup>

Examination of the protein–protein interaction interface has implicated certain regions of the Trx fold in substrate recognition. In Trx-h2 these regions involve residues 45 to 48 (sequence WCGP), 87 to 89 (AMP) and 104 to 106 (VGA). Of particular importance is the residue N-terminal to the *cis*-proline (residue Met 88 in Trx-h2) which forms two backbone hydrogen-bonds with the substrate (see Fig. 4C).<sup>60</sup> Maeda *et al.* suggested



substrate disulfides need to fulfil three criteria to be optimal targets for proteins of the Trx-fold: (1) the backbone amino and carbonyl groups of the target cysteine (*i*) and the carbonyl group of residue *i*-2 should be solvent exposed and free of intermolecular contacts (such as backbone-backbone hydrogen bonds with other secondary structure elements) to allow the formation of backbone hydrogen bonds with a Trx-like oxidoreductase; (2) the peptide chain between the cysteine and the *i*-2 residue should be in extended conformation; and (3) the S<sub>γ</sub> atom of the cysteine residue must be solvent exposed in order to receive the nucleophilic attack from the catalytic Cys residue (residue 46 in Trx-h2).<sup>60</sup> These criteria are fulfilled by most aCSDNs in edge strand contexts. Proteins which contain redox-active aCSDNs that are known substrates of Trx-like proteins include DsbD, glutathione peroxidase, archaeal peroxiredoxins and bacterioferritin comigratory protein. In addition, there are other aCSDNs in our dataset which have been shown to be reduced by Trx, however the biological relevance, if any, of this reactivity has not yet been elucidated. Examples include the barley  $\alpha$ -amylase/subtilisin inhibitor (BASI) aCSDn (note: this disulfide is a member of the Kunitz (STI) inhibitor cluster), and a true-aCSDn-2 linking Cys 106 to Cys 119 of  $\beta$ -lactoglobulin (PDB: 1beb).

It is likely that there are other physiologically redox-active aCSDNs within our dataset which have not yet been identified or investigated. We note that there are several examples in the dataset with crystallographic evidence of redox activity, *i.e.* where an aCSDn is reduced in some structures and oxidised in others (see Tables 3 and S1<sup>†</sup>), however there has been no further investigation into whether this is biologically relevant. Some examples include the disulfide between Cys 181 and Cys 196 in methylamine dehydrogenase H-chain, which is oxidised in PDB: 2j56 and reduced in PDB: 1mg3 (amine dehydrogenase cluster); and an aCSDn in BLIP which connects Cys 30 to Cys 42 in PDB: 3gmu but is reduced between Cys 1030 and Cys 1042 in chains C and D of PDB: 2b5r.

It is hoped that the data in Table S1<sup>†</sup> will be of value to other scientists seeking to identify potential functional sites in proteins for which the function and/or mechanism is as yet unknown.

## D Evolution of aCSDNs

One important question regarding aCSDNs, and indeed all biologically active disulfides, is how they were originally introduced into proteins. Previously we suggested that redox-active disulfides in peroxiredoxins may have been introduced in a stepwise fashion, with the redox-sensitive Cys being introduced first, conferring antioxidant functionality on the protein, and the second Cys being acquired later to provide a regulatory mechanism for the activity.<sup>11</sup> Alternatively, both Cys residues could be acquired simultaneously *via* a retrotransposon. This is a process whereby an existing sequence of DNA is inserted elsewhere in the genome, including into existing genes, resulting in an insertion mutation. If such a retrotransposon already contains an aCSDn, the motif, and hence its functionality, may be added to an existing protein.

Evidence from our dataset indicates that aCSDNs have predominantly been introduced *via* sequential mutation of existing residues to Cys, rather than by retrotransposon insertion. For four proteins in the dataset, intermediate sequences could be found in early diverging organisms that retained a single Cys residue. These have been published in detail elsewhere.<sup>32</sup> For the most part, however, only two states could be discerned: the primordial state without Cys residues and the redox-sensitive state with both Cys present. The clades containing the redox-sensitive state are listed in the taxonomic conservation column of Tables 3 and S1.<sup>†</sup> It is important to note that the sequences available for comparison do not represent the entire suite of species; the lack of an intermediate state may simply be a selection effect and “missing link” sequences bearing a single Cys may yet be found. Alternatively, the single Cys state may be less functionally useful than the aCSDn and thus the intermediate state may only be retained in select instances.

## Conclusions

Here we have investigated the population, embeddedness, disulfide and backbone conformations, and torsional strain of aCSDNs. These include true-aCSDNs embedded in regular  $\beta$ -sheets, which have been described before,<sup>9-11,25,26</sup> a new class of true-aCSDNs, found adjacent to A-bent bulges, as well as aCSDNs found at the termini of  $\beta$ -sheets (end-aCSDNs) and within isolated NHB site  $\beta$ -bridges ( $\beta$ -bridge-aCSDNs). True-aCSDNs are by far the most common aCSDn sub-motifs, being represented by 157 unique disulfide clusters. End-aCSDNs are less common, having only 44 clusters, while there are only 9 clusters of  $\beta$ -bridge-aCSDNs. Several additional clusters have representatives with differing levels of LonE. A disulfide on the edge of a  $\beta$ -sheet is the most common level of LatE although true- and end-aCSDNs have significant populations of all LatE levels.  $\beta$ -Bridge-aCSDNs, on the other hand, show a preference for  $\beta$ -ribbon structures (*i.e.* LonE is correlated with LatE for these structures).

An important outcome of this work is the discovery of the source of the strong preference of all aCSDn sub-motifs for the right-handed staple disulfide conformation. This preference can be traced to the interaction between the disulfide conformation and the inherent right-handed twist of  $\beta$ -sheets. Our *ab initio* calculations revealed that a right-handed disulfide conformation gives rise to a right-handed twist between the two Cys residues which is in excellent agreement with the twist produced by the preferred backbone  $\phi$  and  $\psi$  dihedral angles in protein  $\beta$ -sheets. In contrast, a left-handed conformation gives rise to a twist of the same magnitude but opposite sign (a left-handed twist). Thus the right-handed conformation can fit easily into a  $\beta$ -sheet without producing any additional strain on the structure, while a left-handed aCSDn can only be accommodated by either twisting the sheet away from its natural right-handed curve or by straining the disulfide. We have also identified a second, higher energy disulfide conformation which can be adopted by aCSDNs, the *cis* conformation.



An unexpected outcome of this study is the discovery that end-aCSDns experience, on average, more torsional strain than their true-aCSDn analogues. This was surprising as end-aCSDns experience fewer conformational constraints due to the reduced hydrogen bonding in the disulfide region. We had therefore expected that end-aCSDns should be able to relax their disulfide conformation to reduce the torsional strain. As no such relaxation is seen, it may be inferred that the disulfide conformation is not placing undue stress on the protein system. This, together with the lack of twisting and shearing strain seen for right-handed staple aCSDn motifs, indicates that insertion of an aCSDn into a  $\beta$ -sheet may result in far less strain in the system than originally thought.

Cross-strand disulfides were initially identified as candidates for redox activity due to their relatively high torsional energies and location in protein contexts deemed forbidden.<sup>25</sup> Further studies,<sup>11,44</sup> including this work, have revealed that aCSDns, the most common type of CSD, actually have intermediate disulfide torsional energies. aCSDns are generally less strained than disulfides adopting other forbidden disulfide motifs, or other highly strained disulfides that have been shown to be reduced by mechanical means; however, their torsional energies are usually significantly higher than those of the spiral conformations adopted by most redox-inert, structurally stabilising disulfides (as low as DTE = 0 kJ mol<sup>-1</sup>). This work has also shown that the most common aCSDn conformation, the right-handed staple, is sufficiently compatible with normal  $\beta$ -sheet twist that it can be accommodated with very little strain on the protein structure. This unique property of disulfide metastability, conferring redox-activity at low energetic cost to the overall protein structure, may account for the relative abundance of aCSDns compared to other forbidden disulfide types. aCSDn metastability enhances their role as environmentally regulated biological redox switches, allowing them to perform critical biological functions including signalling responses to physiological redox changes (*e.g.* inflammation), controlling cross-membrane transport and protecting against oxidative stress. As some aCSDns are recognized as specific substrates of thioredoxin-like enzymes, we propose they are cognate substrates of these enzymes. As such, the evolutionary introduction of the motif into non-homologous proteins allows exaptation of the protein into existing thiol-based redox-regulated systems, including redox homeostasis and signalling pathways.

## Methods

### A Construction of reference database and non-redundant dataset

The construction of our reference database of disulfide bonds from the PDB and the clustering of homologous disulfides in BSD motifs to form a non-redundant dataset has been described in detail elsewhere.<sup>11</sup> In brief, the database contains details of all disulfide bonds in X-ray crystal structures with resolutions of 3 Å or better from the PDB Archive Version 4.0 (July 2011). The final dataset consists of 71 337 disulfides from 29 261 PDB files. Roughly 10% of these disulfides (7442) connect adjacent strands of regular  $\beta$ -sheets, with 5187 being

CSDs. These were further split into 4861 aCSDns, 114 aCSDhs and 212 pCSDs. Custom programs were used to identify the disulfides adopting these BSD motifs within the database and to incorporate information on secondary structure motifs from the SCOP database.<sup>64</sup> The torsional strain on the disulfide bond, as well as the twist and shear between the Cys residues were also predicted, as described below.

To reduce bias from multiple structures of the same or similar proteins, we clustered homologous disulfides to give a second, non-redundant, dataset. To be classed as redundant, we required disulfides to both be found within the same protein fold and, critically, to adopt exactly the same position within that fold. Where a protein had already been assigned to a SCOP family, this, together with the Cys residue numbers and the inter-residue distance, was used to identify both homologous disulfides in different proteins and non-homologous disulfides in the same PDB file. Disulfides in structures not assigned to a SCOP family were grouped with the appropriate cluster using the protein structure comparison service "Fold at European Bioinformatics Institute" (<http://www.ebi.ac.uk/msd-srv/ssm>)<sup>65</sup> and the 3D similarity function from the RCSB PDB website (based on the jFATCAT-rigid algorithm),<sup>66</sup> as well as fold data from the InterPro<sup>67</sup> and Pfam<sup>68</sup> databases. Swiss-PDBViewer<sup>42</sup> was used to superimpose and view structures to quality check the clustering process.

### B *Ab initio* calculations

(a) **Estimation of torsional strain in disulfide bonds.** The relative stabilities of disulfide bonds were interpolated from a potential energy surface (PES) for torsion around  $\chi_2$ ,  $\chi_3$  and  $\chi_2'$  (10° resolution). The PES was calculated using the model compound diethyl disulfide at the M05-2X<sup>69</sup>/6-31G(d)<sup>70</sup> level of computational theory with complete geometry optimisations for each data point. Zero point energies were not included as these are too computationally expensive to calculate for each data point. All energies are quoted relative to the lowest energy minimum on the PES; this corresponded to disulfides adopting the spiral conformation. The resulting PES, seen in Fig. 5B and 6B, shows how the torsional strain of the disulfide changes with rotation around  $\chi_2$  (C $_{\alpha}$ C $_{\beta}$ S $_{\gamma}$ S $_{\gamma}'$ ) and  $\chi_2'$  (S $_{\gamma}$ S $_{\gamma}'$ C $_{\beta}$ C $_{\alpha}'$ ). This work has been published in detail elsewhere.<sup>22,44</sup> The  $r(C_{\alpha}-C_{\alpha})$  and C $_{\beta}$ -C $_{\alpha}$ -C $_{\alpha}$ -C $_{\beta}$  dihedral angle contour maps of Fig. 5C and D were also generated from results of these calculations.

Linear interpolation of the points on the PES was used to predict the torsional strain in  $\chi_2$ ,  $\chi_3$ , and  $\chi_2'$  experienced by each disulfide structure in the PDB. Throughout the analysis this is referred to as the disulfide torsional energy (DTE). The effects of strain in  $\chi_1$  and  $\chi_1'$  were estimated with empirical terms from the Amber force field.<sup>71</sup> The sum of the disulfide strain/energy and the  $\chi_1$ ,  $\chi_1'$  strain is referred to as the total torsional energy (TTE).

### (b) Disulfide conformations

The  $\chi_2$  and  $\chi_2'$  dihedral angles (the disulfide conformations) adopted by left-handed and right-handed aCSDns have been plotted separately in Fig. 6A and B. The axes are reversed for



left-handed disulfides so that the same conformations appear in the same regions of the plots. For right-handed aCSDNs the conformational plot was superimposed on a slice of the diethyl disulfide PES. The PES slice used in the figures shows  $\chi_3$  ( $C_\beta S_\gamma S_\gamma' C_\beta'$ ) =  $\pm 100^\circ$ . Generally, a disulfide is most stable when  $\chi_3$  adopts values close to either  $\pm 90^\circ$ , however, for over 95% of right-handed aCSDNs,  $\chi_3$  is found in the range  $85^\circ < |\chi_3| < 115^\circ$ . Hence, the PES corresponding to  $\chi_3 = \pm 100^\circ$  is shown rather than that for the “optimal”  $\chi_3$  value of  $\pm 90^\circ$ . When  $\chi_3$  is outside the range of  $85^\circ$  to  $115^\circ$  the diethyl disulfide PES is sufficiently different that projecting disulfides with these  $\chi_3$  dihedral angles onto the surface is potentially misleading. Thus left-handed aCSDn motifs, for which a significant proportion of disulfides had  $\chi_3$  outside these limits, are not plotted with the PES underlay.

### (c) Investigations of aCSDn handedness preferences

aCSDNs demonstrate a strong preference for right-handed disulfide conformations. In order to investigate the reasons for this preference, M05-2X/6-31G(d) calculations were performed on model compounds for the protein disulfide bond. This model was designed to include all atoms of the Cys residues plus the carbonyl and  $C_\alpha$  of the previous residue, as well as the N and  $C_\alpha$  of the subsequent residue in each protein chain, *i.e.* the peptide linkages before and after each Cys. The  $C_\alpha$  atoms of these additional residues were substituted by methyl groups to terminate the chains. This model allowed one hydrogen bond to be formed on each side of the disulfide, giving a reasonable small scale model of an aCSDn (Fig. 8). This model compound was fully optimised for both left- and right-handed disulfide conformations.

All computational chemistry calculations were performed using the Gaussian03 suite of programs<sup>72</sup> on the AC and XE clusters at the NCI National Facility.

### C Sheet twist and shear

Insertion of a BSD into a  $\beta$ -sheet often results in distortion of the surrounding  $\beta$  structure. In order to obtain a measure of this distortion, the relative twist and shear between  $\beta$ -partners in the region of the disulfide was calculated. Following the method of Ho and Curmi,<sup>47</sup> each residue was described by a vector (the residue backbone vector) connecting the midpoints of its two peptide bonds. (See Fig. S4.†) A strand pair backbone vector could thus be defined as the difference (in the case of antiparallel sheet) or sum (for parallel sheet) of the two residue vectors. The relative shear of two  $\beta$ -partners was calculated by projecting the positions of the  $C_\alpha$  atoms of each residue onto the strand pair backbone vector and finding the distance between these projected points. For the purposes of our work, it was necessary to modify the method of Ho and Curmi for determining the relative twist between two  $\beta$ -partners: instead of using the angle between the two residue backbone vectors, we have used the dihedral angle between them, with the vector connecting the midpoints of the residue backbone vectors being the axis of the dihedral angle. Ho and Curmi were interested in residue pairs within a  $\beta$ -sheet for which the two methods should give

comparable results; we are additionally interested in obtaining a measure of the twist between residues at the end of  $\beta$  ladders, where the two chains may begin to diverge. The use of the dihedral angle reduces the influence of this divergence on the calculated twist.

## Acknowledgements

The authors would like to thank the NCI national facility for a grant of computing time which allowed the calculations mentioned in this paper to be performed.

## References

- 1 B. B. Buchanan, *Annu. Rev. Plant Biol.*, 1980, **31**, 341–374.
- 2 B. B. Buchanan and Y. Balmer, *Annu. Rev. Plant Biol.*, 2005, **56**, 187–220.
- 3 M. A. Wouters, S. W. Fan and N. L. Haworth, *Antioxid. Redox Signaling*, 2010, **12**, 53–91.
- 4 Y. Yang, Y. Song and J. Loscalzo, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10813–10817.
- 5 K. M. Humphries, P. A. Szweida and L. I. Szweida, *Free Radical Res.*, 2006, **40**, 1239–1243.
- 6 G. H. Snyder, M. J. Cennerazzo, A. J. Karalis and D. Locey, *Biochemistry*, 1981, **20**, 6509–6519.
- 7 T. E. Creighton, *J. Mol. Biol.*, 1975, **96**, 767–776.
- 8 J. S. Richardson, *Adv. Protein Chem.*, 1981, **34**, 167–339.
- 9 J. M. Thornton, *J. Mol. Biol.*, 1981, **151**, 261–287.
- 10 M. A. Wouters, R. A. George and N. L. Haworth, *Curr. Protein Pept. Sci.*, 2007, **8**, 484–495.
- 11 N. L. Haworth and M. A. Wouters, *RSC Adv.*, 2013, **3**, 24680–24705.
- 12 C. W. Goulding, M. R. Sawaya, A. Parseghian, V. Lim, D. Eisenberg and D. Missiakas, *Biochemistry*, 2002, **41**, 6920–6927.
- 13 L. J. Matthias, P. T. W. Yam, X. M. Jiang, N. Vandegraaff, P. Li, P. Pountourios, N. Donoghue and P. J. Hogg, *Nat. Immunol.*, 2002, **3**, 727–732.
- 14 J. Ahamed, H. H. Versteeg, M. Kerver, V. M. Chen, B. M. Mueller, P. J. Hogg and W. Ruf, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 13932–13937.
- 15 M. Eriksson, U. Uhlin, S. Ramaswamy, M. Ekberg, K. Regnström, B.-M. Sjöberg and H. Eklund, *Structure*, 1997, **5**, 1077–1092.
- 16 R. A. Reynolds, A. W. Yem, C. L. Wolfe, M. R. Deibel, C. G. Chidester and K. D. Watenpaugh, *J. Mol. Biol.*, 1999, **293**, 559–568.
- 17 N. Ueyama and T. Araki, *J. Am. Chem. Soc.*, 1978, **100**, 4603–4605.
- 18 N. Srinivasan, R. Sowdhagini, C. Ramakrishnan and P. Balaram, *Int. J. Pept. Protein Res.*, 1990, **36**, 147–155.
- 19 M. A. Wouters and P. M. G. Curmi, *Proteins*, 1995, **22**, 119–131.
- 20 K. Gunasekaran, C. Ramakrishnan and P. Balaram, *Protein Eng.*, 1997, **10**, 1131–1141.
- 21 E. G. Hutchinson, R. B. Sessions, J. M. Thornton and D. N. Woolfson, *Protein Sci.*, 1998, **7**, 2287–2300.



- 22 N. L. Haworth, J. Y. Liu, S. W. Fan, J. E. Gready and M. A. Wouters, *Aust. J. Chem.*, 2010, **63**, 379–387.
- 23 S. Indu, V. Kochat, S. Thakurela, C. Ramakrishnan and R. Varadarajan, *Proteins*, 2011, **79**, 244–260.
- 24 K. Chakraborty, S. Thakurela, R. S. Prajapati, S. Indu, P. S. S. Ali, C. Ramakrishnan and R. Varadarajan, *Biochemistry*, 2005, **44**, 14638–14646.
- 25 M. A. Wouters, K. K. Lau and P. J. Hogg, *BioEssays*, 2004, **26**, 73–79.
- 26 N. L. Haworth, L. L. Feng and M. A. Wouters, *J. Bioinf. Comput. Biol.*, 2006, **4**, 155–168.
- 27 P. M. Harrison and M. J. E. Sternberg, *J. Mol. Biol.*, 1996, **264**, 603–623.
- 28 P. J. Kraulis, *J. Appl. Crystallogr.*, 1991, **24**, 946–950.
- 29 E. A. Merritt and D. J. Bacon, in *Methods in Enzymology*, Academic Press Inc., San Diego, 1997, vol. 277, pp. 505–524.
- 30 N. L. Haworth and M. A. Wouters, 2015, in preparation.
- 31 N. L. Haworth and M. A. Wouters, 2015, in preparation.
- 32 K. A. Mohanasundaram, N. L. Haworth, M. P. Grover, T. M. Crowley, A. Goscinski and M. A. Wouters, *Front. Pharmacol.*, 2015, **6**, 1, DOI: 10.3389/fphar.2015.00001.
- 33 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.
- 34 G. Gopalan, Z. Y. He, K. P. Battaile, S. Luan and K. Swaminathan, *Proteins*, 2006, **65**, 789–795.
- 35 N. Cerutti, B. V. Mendelow, G. B. Napier, M. A. Papathanasopoulos, M. Killick, M. Khati, W. Stevens and A. Capovilla, *J. Biol. Chem.*, 2010, **285**, 25743–25752.
- 36 E. van Anken, R. W. Sanders, I. M. Liscaljet, A. Land, I. Bontjer, S. Tillemans, A. A. Nabatov, W. A. Paxton, B. Berkhout and I. Braakman, *Mol. Biol. Cell*, 2008, **19**, 4298–4309.
- 37 J. Melchers, M. Diechtierow, K. Feher, I. Sinning, I. Tews, R. L. Krauth-Siegel and C. Muhle-Goll, *J. Biol. Chem.*, 2008, **283**, 30401–30411.
- 38 F. F. de Rezende, A. Martins Lima, S. Niland, I. Wittig, H. Heide, K. Schröder and J. A. Eble, *Free Radicals Biol. Med.*, 2012, **53**, 521–531.
- 39 S. W. Fan, R. A. George, N. L. Haworth, L. L. Feng, J. Y. Liu and M. A. Wouters, *Protein Sci.*, 2009, **18**, 1745–1765.
- 40 J. J. Wey, S. S. Tang and T. Y. Wu, *Acta Pharmacol. Sin.*, 2006, **27**, 1238–1246.
- 41 B. Pillai, M. M. Cherney, K. Hiraga, K. Takada, K. Oda and M. N. G. James, *J. Mol. Biol.*, 2007, **365**, 343–361.
- 42 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.
- 43 V. P. ZavYalov, T. V. Chernovskaya, D. A. G. Chapman, A. V. Karlyshev, S. MacIntyre, A. V. Zavialov, A. M. Vasiliev, A. I. Denesyuk, G. A. ZavYalova, I. V. Dudich, T. Korpela and V. M. Abramov, *Biochem. J.*, 1997, **324**, 571–578.
- 44 N. L. Haworth, J. E. Gready, R. A. George and M. A. Wouters, *Mol. Simul.*, 2007, **33**, 475–485.
- 45 E. Mizohata, H. Sakai, E. Fusatomi, T. Terada, K. Murayama, M. Shirouzu and S. Yokoyama, *J. Mol. Biol.*, 2005, **354**, 317–329.
- 46 T. Nakamura, Y. Kado, T. Yamaguchi, H. Matsumura, K. Ishikawa and T. Inoue, *J. Biochem.*, 2010, **147**, 109–115.
- 47 B. K. Ho and P. M. G. Curmi, *J. Mol. Biol.*, 2002, **317**, 291–308.
- 48 S. J. Liao, C. Y. Yang, K. H. Chin, A. H. J. Wang and S. H. Chou, *J. Mol. Biol.*, 2009, **390**, 951–966.
- 49 U. Krug, M. Zebisch, M. Krauss and N. Sträter, *J. Biol. Chem.*, 2012, **287**, 3051–3066.
- 50 A. Weichsel, J. R. Gasdaska, G. Powis and W. R. Montfort, *Structure*, 1996, **4**, 735–751.
- 51 A. Zapun, S. Grammatyka, G. Deral and T. Vernet, *Biochem. J.*, 2000, **350**, 873–881.
- 52 M. F. López-Lucendo, D. Solís, J. L. Sáiz, H. Kaltner, R. Russwurm, S. André, H.-J. Gabius and A. Romero, *J. Mol. Biol.*, 2009, **386**, 366–378.
- 53 M. Kim, H. J. Min, H. Y. Won, H. Park, J.-C. Lee, H.-W. Park, J. Chung, E. S. Hwang and K. Lee, *PLoS One*, 2009, **4**, e6464.
- 54 M. Kim, J. Maeng and K. Lee, *Biochimie*, 2013, **95**, 659–666.
- 55 J. J. Calvete, F. Revert, M. Blanco, J. Cervera, C. Tarrega, L. Sanz, F. Revert-Ros, F. Granero, E. Perez-Paya, B. G. Hudson and J. Saus, *Proteomics*, 2006, **6**, S237–S244.
- 56 M. E. Kimple, D. P. Siderovski and J. Sondek, *EMBO J.*, 2001, **20**, 4414–4422.
- 57 K. Reiser, K. O. François, D. Schols, T. Bergman, H. Jörnvall, J. Balzarini, A. Karlsson and M. Lundberg, *Int. J. Biochem. Cell Biol.*, 2012, **44**, 556–562.
- 58 H. Ostergaard, A. Henriksen, F. G. Hansen and J. R. Winther, *EMBO J.*, 2001, **20**, 5853–5862.
- 59 R. E. Hansen, H. Ostergaard and J. R. Winther, *Biochemistry*, 2005, **44**, 5899–5906.
- 60 K. Maeda, P. Häggglund, C. Finnie, B. Svensson and A. Henriksen, *Structure*, 2006, **14**, 1701–1710.
- 61 G. del Val, B. C. Yee, R. M. Lozano, B. B. Buchanan, R. W. Ermel, Y.-M. Lee and O. L. Frick, *J. Allergy Clin. Immunol.*, 1999, **103**, 690–697.
- 62 Z. X. Zhang, D. Y. Geng, Q. Han, S. D. Liang and H. R. Guo, *J. Fish Biol.*, 2013, **83**, 1287–1301.
- 63 K. Maeda, C. Finnie and B. Svensson, *Biochem. J.*, 2004, **378**, 497–507.
- 64 A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia, *J. Mol. Biol.*, 1995, **247**, 536–540.
- 65 E. Krissinel and K. Henrick, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 2256–2268.
- 66 Y. Ye and A. Godzik, *Bioinformatics*, 2003, **19**, ii246–ii255.
- 67 S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. A. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajananathan, P. D. Thomas, C. H. Wu, C. Yeats and S.-Y. Yong, *Nucleic Acids Res.*, 2012, **40**, D306–D312.
- 68 M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements,



- A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman and R. D. Finn, *Nucleic Acids Res.*, 2012, **40**, D290–D301.
- 69 Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 364–382.
- 70 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
- 71 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1–41.
- 72 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision E.01*, Gaussian, Inc., Wallingford CT, 2004.

