

Computational survey of peptides derived from disulphide-bonded protein loops that may serve as mediators of protein-protein interactions.

AUTHOR(S)

Fergal J. Duffy, Marc Devocelle, David R. Croucher, Denis C. Shields

CITATION

Duffy, Fergal J.; Devocelle, Marc; Croucher, David R.; Shields, Denis C. (2014): Computational survey of peptides derived from disulphide-bonded protein loops that may serve as mediators of protein-protein interactions.. Royal College of Surgeons in Ireland. Journal contribution.
<https://hdl.handle.net/10779/rcsi.10991465.v2>

HANDLE

[10779/rcsi.10991465.v2](https://hdl.handle.net/10779/rcsi.10991465.v2)

LICENCE

CC BY-NC-ND 4.0

This work is made available under the above open licence by RCSI and has been printed from <https://repository.rcsi.com>. For more information please contact repository@rcsi.com

URL

https://repository.rcsi.com/articles/journal_contribution/Computational_survey_of_peptides_derived_from_disulphide-bonded_protein_loops_that_may_serve_as_mediators_of_protein-protein_interactions_/10991465/2

RESEARCH ARTICLE

Open Access

Computational survey of peptides derived from disulphide-bonded protein loops that may serve as mediators of protein-protein interactions

Fergal J Duffy^{1,2,3}, Marc Devocelle⁵, David R Croucher⁴ and Denis C Shields^{1,2,3*}

Abstract

Background: Bioactive cyclic peptides derived from natural sources are well studied, particularly those derived from non-ribosomal synthetases in fungi or bacteria. Ribosomally synthesised bioactive disulphide-bonded loops represent a large, naturally enriched library of potential bioactive compounds, worthy of systematic investigation.

Results: We examined the distribution of short cyclic loops on the surface of a large number of proteins, especially membrane or extracellular proteins. Available three-dimensional structures highlighted a number of disulphide-bonded loops responsible for the majority of the likely binding interactions in a variety of protein complexes, due to their location at protein-protein interfaces. We find that disulphide-bonded loops at protein-protein interfaces may, but do not necessarily, show biological activity independent of their parent protein. Examining the conservation of short disulphide bonded loops in proteins, we find a small but significant increase in conservation inside these loops compared to surrounding residues. We identify a subset of these loops that exhibit a high relative conservation, particularly among peptide hormones.

Conclusions: We conclude that short disulphide-bonded loops are found in a wide variety of biological interactions. They may retain biological activity outside their parent proteins. Such structurally independent peptides may be useful as biologically active templates for the development of novel modulators of protein-protein interactions.

Keywords: Cyclic peptide, Protein loop, Protein interface, Bioactive peptide, Ribosomal cyclic peptide

Background

Cyclic peptides are macrocyclic peptides which possess where linear peptide side chains or termini are covalently bonded to shape the peptide into a ring. Macrocyclic compounds such as cyclic peptides have been a renewed focus of drug discovery in recent years [1], and identifying a biologically pre-designed set of cyclic peptides in protein sequences would be of great potential interest in pharmaceutical development. Many cyclic peptides

are disulphide-cyclic, where two cysteine residues form an S-S bond between two thiol side chain groups to cyclise the peptide. Many other types of cyclic peptide are possible, including head-tail cyclised peptides, where the amino N-terminus and carboxy C-terminus are bound together with an amide bond to cyclise the peptide backbone, and other side-chain crosslinked cyclic peptides, such as those possessing amide bond between a lysine side chain and an aspartic acid side chain. Side-chain crosslinked cyclic peptides include disulphide-cyclic peptides, such as those described in this work; cyclic peptides where the side-chains mimic a peptidic bond, such as Lysine-Aspartic acid side chain cyclised peptides; as well as depsipeptides, where the side-chain crosslinking bond is an ester bond. An example cyclic depsipeptide

*Correspondence: denis.shields@ucd.ie

¹School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland

²Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland

Full list of author information is available at the end of the article

would be a serine-aspartic acid side-chain cyclised peptide. There also exist backbone head-tail cyclic peptides with additional disulphide bonding such as the cyclotides [2] and theta-defensins [3], where the disulphide bonding serves to provide additional conformational constraint to the already cyclic peptide. Cyclic peptides are interesting from a drug development point of view, due to their generally better specificity, proteolytic resistance, and stability than linear peptides [4]. While the term “cyclic peptide” can refer to a peptide cyclised by any of the above strategies, the term is often used to refer to head-tail cyclised peptides. Therefore, this work use the term “small macrocyclic peptides” to refer to cyclic peptides in general, including the disulphide-cyclised peptides described in this work.

Many examples of bioactive and therapeutic natural peptides are known, including antibiotics [5], natural hormone mimics [6] and immunosuppressants [7]. The source of these peptides are mostly non-ribosomal plant and fungal secondary metabolites, produced by specialised non-ribosomal peptide-synthetase enzymes. These can create peptides with a wide range of unusual amino acids, that can be of varying chirality, carry modified side chains and backbones, and be cyclised [8]. Reviews by Conlon et al. [9] and Cascales et al. [10] provide more detailed overview of the diversity of small macrocyclic peptides in nature. The Cybase [11,12] database has been developed as a publically available resource describing of sequence, structural and functional properties of naturally derived cyclic proteins and peptides.

The biological activity of naturally occurring ribosomally synthesised peptides have not been subject to any systematic surveys. Disulphide bonds are known to play a key structural role in proteins, stabilising the protein tertiary structure on a large scale, and can also influence quaternary structure via interchain disulphide bonds. However, disulphide bonds can also have a local effect in constraining a much smaller component of the structure. A protein loop is a general term for a protein secondary structural element which is not a helix or sheet, and generally exhibit a lack of hydrogen bonding and high flexibility, and often served to join other secondary structural elements. Protein turns are specific types of loops where the polypeptide chain reverses its overall direction and can be between 1 and 5 residues long (α , β , γ , δ and π turns) [13]. A special case of this is the β -hairpin turn, which connects two antiparallel β -sheets. These regions are known to be important in protein-protein contacts [14], and short (2-8 amino-acid) protein loops or turns can be “pinned” in place by a disulphide bond [15], forming a surface structural motif held in a relatively fixed position by the disulphide bond, thus having a certain amount of independence from the larger protein tertiary structure. This

approach has previously been explored in phage-display studies [16].

The idea of finding “self-inhibitory” peptides, where a peptide derived from a protein-protein interface inhibits the formation of that interface, has been previously explored, and it has been found that many of globular protein interactions are dominated by linear peptide segments [17]. Among other applications, this approach has recently been used to identify peptides that inhibit viral membrane fusion [18]. Therefore, in principle short disulphide-bonded loops derived from protein sequences and located at protein interfaces could be synthesised separately and show a similar or related biological activity to the parent protein.

The use of small macrocyclic peptides to mimic protein loops has also been exploited for the RGD peptides [19]. The RGD tripeptide motif is a cell attachment β -turn motif found in numerous proteins, and small macrocyclic peptides containing this motif have been shown to inhibit integrin $\alpha V \beta 3$ activity, which plays an important role in tumour metastasis.

Traditionally, protein-protein interaction inhibitors are discovered by screening compounds against a particular known “target” interaction of interest in a biochemical pathway. Our motivation in this study is to harness the vast amount of protein sequence and structural data available to develop a bioinformatic approach to identifying candidate bioactive small macrocyclic peptides from disulphide-bonded protein loops. In contrast to screening compounds against a single target, this method of analysis allows simultaneous identification of modulators against a variety of protein-protein interactions that have been evolutionarily selected for. This type of bioinformatics approach has been previously successfully used to identify peptides from signalling rich juxtamembrane regions that have the ability to modulate platelet function [20]. In this study, we surveyed the sequence, structural and conservation properties of disulphide bonded protein loops, in order to infer a set of small macrocyclic peptides capable of bioactivity outside the context of their parent protein.

Results and discussion

Finding disulphide-bonded loops at protein-protein interfaces

To identify short disulphide bonded loops that play a crucial role at Protein-Protein Interfaces, we set out to find known three-dimensional structures of protein complexes mediated by a disulphide bonded loops. We defined short disulphide bonded loops as those equal to or less than eleven residues in length (two flanking cysteines, plus 2-9 internal residues), and excluding cysteine-knot like regions of overlapping disulphide bridges.

2,380 Protein Data Bank (PDB) structures [21] corresponding to Uniprot entries with annotated short

disulphide-bonded loops were downloaded. (Figure 1(a)) Disulphide-bonded loops at protein surfaces or interfaces were identified as described in the Methods. The number of surface residues (defined as those having at least 2.5\AA^2 of solvent exposed surface area), and interface residues (those within 3\AA of another protein chain) were counted using PyMol [22]. After removing redundant PDB structures, we found 1,231 short disulphide-bonded loops at protein surfaces (Figure 1(b)), and 132 short disulphide-bonded loops located at known protein-protein interfaces (Figure 1(c)). Of these 132 disulphide-bonded loops, 13 occupied over 50% of the interface surface area. We also found that the number of protein-protein interface hotspots contained in these disulphide-bonded loops, as predicted by the HotRegion [23] server, is generally proportional to the interface area covered by the loops (Additional file 1: Figure S1). Thus, the majority of disulphide-bonded loops at protein-protein interfaces do not constitute the major determinant of the interaction. Nevertheless, they may still play key roles in binding and represent potential templates for mimetics that target protein-protein interactions.

The proteins containing short disulphide-bonded loops shown in Table 1, are generally membrane, secreted,

or extracellular, with exceptions being palmitoyl-protein thioesterase 1, in the lysosome. The reason for this is likely that the disulphide bonds would not last in the reducing environment of the cytosol. Disulphide bonds can exist in the cytosol, but they are generally either embedded in the hydrophobic core of the protein, protected from the cytosol, or in the presence of sulphhydryl oxidases [24].

Table 1 identifies the 13 proteins containing a short disulphide-bonded loop comprising over 50% of the surface area of a PDB protein-protein interface, along with the partner protein at the interface. These interfaces come from a variety of species, including bacteria (*Alcaligenes sp.*, *E. coli*), plants (*P. angularis*, *V. radiata*, *V. unguiculata*), and animals (*H. sapiens*, *M. musculus*, *N. atra*, *B. irregularis*). They also cover a variety of protein types, including snake venoms, proteinase inhibitors, collagen, and extracellular proteins involved in growth and cell adhesion.

Structural independence of short disulphide loops

In order to explore the structural independence of short disulphide bonded loops when removed from their parent protein, structural models of the short disulphide loops described in Table 2 were generated using the PEP-FOLD

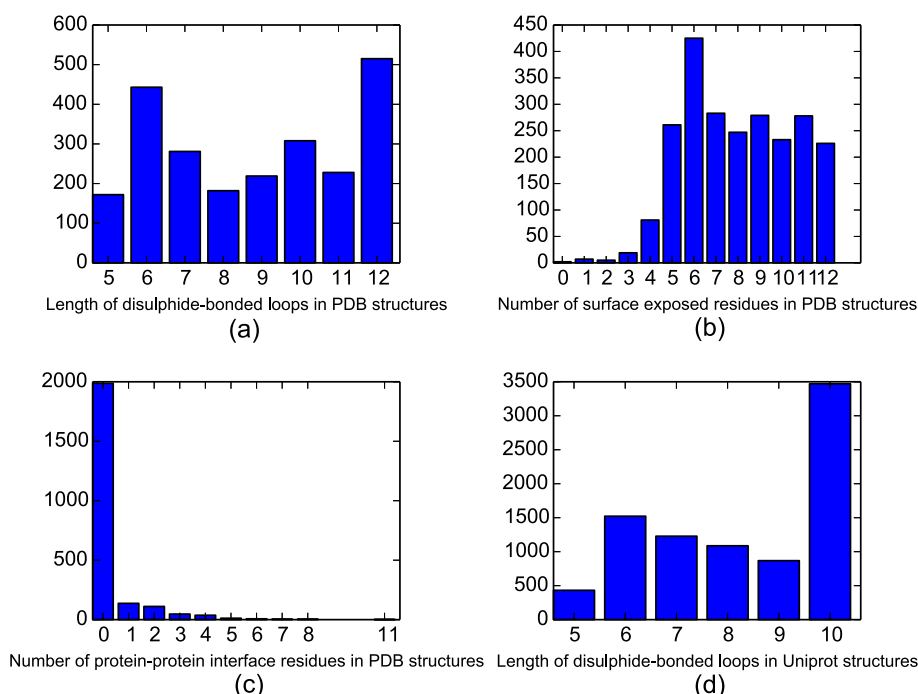


Figure 1 Loop, surface and interface size distributions for disulphide-bonded loop containing proteins. This study examined a total of 2,380 3D loop structures from 759 unique proteins in the Protein Data Bank (PDB), along with 8,607 annotated disulphide-bonded loop sequences from 5,989 Uniprot protein entries. **(a)** Shows the loop size distribution of disulphide bonded loops in PDB structures. **(b)** Shows the distribution of the number of surface residues possessed by each PDB-derived disulphide-bonded loop **(c)** shows the distribution of the number of residues in each PDB-derived disulphide loop that are involved in a protein-protein contact. **(d)** Shows the size distribution of short disulphide-bonded loops annotated in Uniprot.

Table 1 Uniprot Proteins containing a disulphide-bonded loop comprising over 50% of a PDB protein-protein interface

Uniprot accession	Protein containing disulphide-bonded loop	Interacting protein	PDB ID
P02462	Collagen alpha-1(IV) chain [Cleaved into: Arresten], <i>Homo sapiens</i>	Homodimer	1LI1
Q7T1K6	Natrin-1 (Cysteine-rich venom protein 1) (NA-CRVP1) (Protein G2a), <i>Naja atra</i>	Homodimer	1XTA
P01133	Pro-epidermal growth factor (EGF) [Cleaved into: Epidermal growth factor (Urogastrone)] , <i>Homo sapiens</i>	EGF Receptor (ErbB1), <i>Homo sapiens</i>	1IVO
P01058	Bowman-Birk type proteinase inhibitor, <i>Phaseolus angularis</i>	Bovine trypsin	1TAB
P01062	Bowman-Birk type trypsin inhibitor, <i>Vigna radiata</i>	Trypsin, <i>Bos taurus</i>	1G9I
P17734	Bowman-Birk type seed trypsin and chymotrypsin inhibitor (BTCL), <i>Vigna unguiculata</i>	Trypsin, <i>Bos taurus</i>	2G81
P0AD59	Inhibitor of vertebrate lysozyme, <i>Escherichia coli</i>	Lysosome C, <i>Gallus gallus</i>	1GPQ
Q9JHF9	Lymphocyte antigen 96 (Ly-96) (ESOP-1) (Protein MD-2), <i>Mus musculus</i>	Toll-like receptor 4 (TLR-4), <i>Mus musculus</i>	2Z64
A0S865	Iridotoxin subunit B (IrTxB), <i>Boiga irregularis</i>	Iridotoxin subunit A	2H7Z
P50897	Palmitoyl-protein thioesterase 1 (PPT-1) (EC 3.1.2.22) (Palmitoyl-protein hydrolase 1), <i>Homo sapiens</i>	Homodimer	3GRO
P04004	Somatomedin-B subunit of Vitronectin (VN), <i>Homo sapiens</i>	Plasminogen activator inhibitor-1 (PAI-1), <i>Homo sapiens</i>	1OCO
P04004	Somatomedin-B subunit of Vitronectin (VN), <i>Homo sapiens</i>	Urokinase plasminogen activator surface receptor (uPAR), <i>Homo sapiens</i>	3BT2
Q03405	Urokinase plasminogen activator surface receptor (uPAR), <i>Homo sapiens</i>	Urokinase-type plasminogen activator (uPA), <i>Homo sapiens</i>	3BT2
Q8RS40	Beta-1,3-xylanase (txyA), <i>Alcaligenes sp.</i>	Homodimer	2COV

[25] *de-novo* peptide structure prediction webserver. These models were generated from the sequence of the disulphide loop alone. Five PEP-FOLD model structures were generated for each disulphide bonded loop in Table 2. The PyMol [22] align tool was then used to align each model disulphide loop to the PDB crystal structure based on backbone C α atoms, and calculate an RMSD between the crystal structure and model. The complete results are shown in Additional file 1: Table S1.

Table 2 Protein families containing preferentially conserved disulphide-bonded loop

Protein family	Disulphide-bonded loop count
Somatotropin	91
Prolactin	50
Polygalacturonase/Endopolygalacturonase	50
Guanylate cyclase activator/Guanylin	13
Urotensin	12
Disintegrin	8
Calcitonin	6
Other	57
Total	287

(Conservation difference between disulphide-bonded loop and juxtapeptide region of over 0.30.).

The mean RMSD of the lowest energy PEP-FOLD model to each corresponding crystal structure was $2.468 \pm 0.767\text{\AA}$. As a point of reference, the ongoing Critical Assessment of Techniques of Protein Structure Prediction (CASP) experiment describes the generation of a homology structure with an accuracy of better than 6.5\AA as “not trivial”, and models with an accuracy of 1.5\AA as “high-resolution”. Models with an RMSD of 4\AA can be considered as having a broadly correct fold [26]. Using these values as a guideline it seems that disulphide-bonded loop structure prediction using the loop sequence alone is sufficient to predict a moderately accurate structure, lending support to the idea that these loops have a large degree of structural independence from their parent protein.

Short disulphide-bonded loop mediated interfaces

For many of the heterodimeric interfaces listed in Table 1, there is additional experimental evidence to suggest that the short disulphide-bonded loops play a key role in binding. In the case of the interaction of Egg Lysosyme and Inhibitor of Vertebrate Lysosyme (Ivy), a cyclic loop on the surface of Ivy (CKPHDC) has been shown to be essential for its inhibitory effect, as shown by mutagenesis studies [27]. This loop is strictly conserved across 30 members of the Ivy family, and 5 other members contain a related CExxDxC motif.

For the Lymphocyte antigen 96 and Toll-like receptor 4 interaction, a disulphide-bonded loop CHGHDDDYSC sits on the conserved “A” patch of Lymphocyte antigen 96. Mutations of five separate amino acids in this peptide have been shown to disrupt binding, including both cysteine residues, implying that the cyclic nature of the peptide is important for complex formation [28].

A CSYYQSC disulphide-bonded loop contained within Vitronectin is located at the interface of two protein-

protein complexes, those of Vitronectin with Plasminogen activator inhibitor-1 and Vitronectin with Urokinase plasminogen activator receptor. The relevance of the short disulphide-bonded loop to the interaction of Vitronectin and Plasminogen activator inhibitor-1 has been experimentally verified. Alanine scanning shows that deletion of Asp22, Glu23, Leu24, Tyr27 and Tyr28 significantly reduces binding affinity. The cyclic region of the interface covers the Cys25 to Cys31 region, including 2 of the 5 critical residues [29]. The same Vitronectin disulphide-bonded loop is critical in the binding of Vitronectin to urokinase plasminogen activator receptor. The four serine (Ser26, Ser30) and tyrosine (Tyr27, Tyr28) residues in the disulphide-bonded loop fit into a cavity on Urokinase plasminogen activator receptor that shows a high degree of shape and charge complementarity [30]. Notably, the same disulphide-bonded loop tyrosine residues are important in Vitronectin binding to both Urokinase plasminogen activator receptor and Plasminogen activator inhibitor-1. Also, a Urokinase plasminogen activator receptor disulphide-bonded loop, CKTNNGDC, is involved in the interaction of Urokinase plasminogen activator receptor with Urokinase plasminogen activator, in direct contact with the Urokinase plasminogen activator Kringle domain. However, this Urokinase plasminogen activator receptor disulphide-bonded loop is not located close to the Vitronectin binding region [30]. The three proteins Urokinase plasminogen activator receptor, Urokinase plasminogen activator and Vitronectin play important roles in regulating the proteolytic degradation of the extracellular matrix and blood clots [31]. This network can also play a role in cancer progression, including degrading the extracellular matrix to facilitate cancer metastasis [32]. Thus a small macrocyclic peptide that can interfere with this network would be of clinical interest.

The Epidermal growth factor disulphide-bonded loop CVVGYIGERC interacts with domain I and III of the Epidermal growth factor receptor [33]. The position of the disulphide-bonded loop region at the interface is evident in Additional file 1: Figure S2, but there is no evidence whether this disulphide-bonded loop may have independent activity. The Epidermal growth factor receptor is part of the ErbB family of receptor tyrosine kinases which act as receptors for a variety of different Epidermal growth factor-domain containing growth factor ligands. All of these Epidermal growth factor domains contain homologous C-terminal disulphide-bonded loops.

Along with their role at the binding interfaces between different proteins, it can be seen from Table 1 that disulphide-bonded loops are also involved in interactions between protein subunits (NatrIn-1, Iridotoxin), and between homodimeric interfaces (Collagen α -1(IV) homodimer, Palmitoyl-protein thioesterase homodimer).

For the cases discussed above (Inhibitor of Vertebrate Lysosyme, Lymphocyte antigen 96, Vitronectin, Epidermal growth factor), the disulphide-bonded loop contains all or part of the protein region responsible for the interaction, and it would be interesting to determine if the disulphide-bonded loop portion of the interface alone is sufficient to modulate the protein-protein interaction.

A disulphide-bonded loop region in Bowman Birk inhibitors possesses independent activity

For one of the interfaces in Table 1 it has been previously shown that the short disulphide-bonded loop portion of the interface is not only a critical part of the interaction, but can act independently of the parent protein. This is the Bowman-Birk family of serine proteinase inhibitors, where the a single disulphide-bonded loop in the protein can act as an inhibitor of trypsin at nanomolar concentrations. The active disulphide loop of the Bowman-Birk inhibitor is nine residues, and Luckett et al. [34] have identified a natural sunflower cyclic Bowman-Birk inhibitor 14 residues long (SFTI-1).

The inhibitory abilities of the disulphide-bonded loop alone has been demonstrated by Domingo et al. [35], who designed a set of eleven residue small macrocyclic peptide loops based on this loop from a variety of Bowman-Birk serine protease inhibitors, and showed that the resulting small macrocyclic peptides inhibit a similar set of serine proteases as the parent protein. The native Bowman-Birk proteins inhibit at picomolar concentrations, and the disulphide-bonded loops at nanomolar concentrations - which is the range that would be expected from a druglike molecule. This result is promising from the point of view of using these loops as lead peptides for drug discovery efforts.

EGF domain small macrocyclic peptides do not show independent activity

To investigate if human disulphide-bonded loops at PDB interfaces can have biological activity when extracted from their parent protein, we tested whether the EGF disulphide-bonded loop identified in Table 1 along with a panel of 14 homologous disulphide-bonded loops from related EGF-domains (Table 3) could modulate EGF receptor behaviour. These sequences were chosen to represent well-described ErbB agonists [36]: a closely related set of growth factors along with a sampling of other diverse EGF domains. The disulphide-bonded loops listed in Table 3 are mostly derived from a closely related set of growth factors in the ErbB pathway, along with a sampling of other diverse EGF domains. Thus, from a single candidate disulphide-bonded loop, a library of related disulphide-bonded loops generated during evolution can be assessed for potential biologic activity.

Table 3 Cyclic peptides derived from EGF-domain containing proteins and tested for EGF activation/inhibition

No.	Cyclic peptide sequence	Human parent protein
1	CWGYIGERC	EGF, Epidermal Growth Factor*
2	CHSGYVGARC	TGFA, Transforming Growth Factor Alpha*
3	CQQEYFGERC	AREG, Amphiregulin*
4	CDEGYIGARC	BTC, Betacellulin*
5	CHPGYHGERC	HBEGF, Heparin-binding EGF-like growth factor*
6	CEVGYTGVR	EREG, Epregrulin*
7	CQPGFTGARC	NRG1, Neuregulin-1**
8	CPNGFFGQRC	NRG2, Neuregulin-2**
9	CKEGYQGVRC	NRG3, Neuregulin-3**
10	CVENYTGARC	NRG4, Neuregulin-4**
11	CAQECVHGR	PEAR1, Platelet endothelial aggregation receptor 1
12	CTRTGYSGPNC	PTGS1, Prostaglandin G/H synthase 1
13	CDPGFSLKC	SELE, E-selectin
14	CLPAFEGRNC	F7, Coagulation factor VII
15	CDSDWTYGYC	ITGB3, Integrin Beta 3

*Denotes growth factors known to directly activate the EGF receptor. **Denotes growth factors known to directly activate members of the ErbB receptor family other than EGFR/ErbB1.

These loop-derived small macrocyclic peptides were synthesised separately and tested for their ability to 1) activate the EGF receptor and 2) competitively inhibit the EGF receptor in the presence of native EGF. Western blotting was used to assess the amount of phosphorylated EGFR after treatment with EGF, small macrocyclic peptide, or EGF following incubation with small macrocyclic peptide, as a proxy for activation. However none of the selected peptides demonstrated any ability to either activate or inhibit the EGF receptor (Additional file 1: Figure S3). We conclude that these peptides do not show significant biological activity independent of their parent protein, in contrast with the Bowman-Birk protease inhibitor peptides.

It is possible that the disulphide loop takes a significantly different shape when removed from the context of the wider EGF protein, hence explaining the lack of biological activity observed. Additional file 1: Table S1 shows that the lowest energy *de-novo* model of this loop has an RMSD of 2.374Å based on the Cα alignment. This suggests that the free peptide retains a structure reasonably close to what has been seen in the crystal structure.

To explain why these EGF peptides do not have activity, we examined the structure of the EGF-EGFR complex.

(PDB ID: 1IVO). The EGFR protein comprises three structural domains (I, II, and III). EGF activates EGFR by binding to a cavity between EGFR domain I and III, with binding sites existing on both domain I and III [33]. The CVVGYIGERC loop (Cys33 - Cys41 of EGF) tested here comprises a large portion of the total EGF-Domain I interface contacts in the crystal structure, but only a small proportion of the EGF-Domain III contacts (Additional file 1: Figure S2). Residues in the C-terminal end of EGF, such as Leu47 are known to make important contacts with Domain III. Thus, despite comprising a large portion of the interface, the disulphide loop is not able to fill the EGFR cavity on both sides, which would likely explain why the disulphide bonded loop is not able to conformationally shift EGFR to its active position. It is possible that the disulphide bonded loop is binding to Domain I of EGFR, but clearly any potential binding is not strong enough to compete with EGF binding to its native receptor.

Conservation of disulphide-bonded loops

The cyclic-peptide mediated interfaces above represent an interesting set of compounds, but it is also of interest to see if disulphide-bonded loops represent a widely used natural strategy to influence protein-protein interactions, by examining evolutionary conservation of short disulphide-bonded loops in proteins.

A dataset of short disulphide-bonded loop containing proteins was assembled from the SwissProt database of manually annotated proteins. Searching for all SwissProt proteins containing short disulphide bonded loops (annotated intrachain disulphide bonds with 2-8 internal residues) revealed 8607 annotated short disulphide-bonded loops in 5989 proteins (Figure 1(d) shows the size distribution of these loops). Figure 2 illustrates the distribution of amino acids in short disulphide-bonded loops, as compared to that of the full range of proteins in Uniprot. Short disulphide-bonded loops seem to contain fewer hydrophobic residues (Valine, Leucine, Isoleucine, Alanine, Methionine) which could indicate that disulphide-bonded loop loops are relatively unlikely to be located at the hydrophobic core of a protein. There is also an enrichment in Glycine and Proline residues, which are known to enable protein backbone flexibility [37], and break up helical structures [38], which may enable turns, helping the cycle to be formed.

Homologs of SwissProt proteins containing annotated short disulphide-bonded loops were identified using the Gopher [39] webserver (bioware.ucd.ie), searching the default set of model organisms. All short disulphide-bonded loop containing proteins with at least one Gopher-identified ortholog were then aligned using MUSCLE [40]. Per-residue conservation scores were then calculated for each alignment using the Jensen-Shannon divergence method of Capra and Singh [41]. Aligned

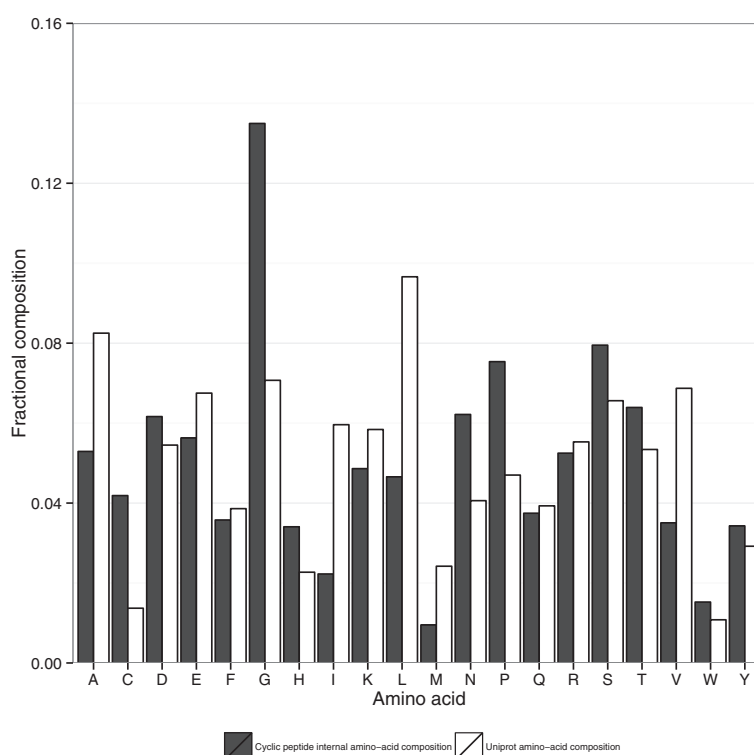


Figure 2 Amino-acid distribution for proteins containing short disulphide-bonded loops. White bars indicate fractional amino acid frequencies across all Uniprot proteins and black bars indicate amino acid frequencies inside short disulphide-bonded loops, excluding the disulphide-bond forming cysteines.

short disulphide regions between the original protein and homolog were identified by examining alignments of the annotated disulphide regions of the original protein. If the loop terminal cysteine residues in the original protein exactly aligned with cysteine residues in the homolog protein, this region was considered a conserved disulphide loop. It is well known that cysteines involved in disulphide bonding are well conserved across a variety of protein families [42]. For this reason, only the conservation of the interior loop residues was considered in this study.

Short disulphide-bonded loops show increased conservation relative to other regions in the same set of proteins (Figure 3). It is possible that the higher conservation of these loops come from loop sequences being dominated by turn-promoting residues such as proline and glycine, which are somewhat over-represented in disulphide bonded loops, as shown in Figure 2. However it has also been observed that residues with high solvent accessibility are more variable between homologs [43,44] than internal residues. Since we have observed that known 3D disulphide bonded loop residues are mostly exposed on the protein surface (Figure 1(b)), this variability will tend to counteract the increased likelihood of proline and glycine residues in the loop aligning by chance.

Disulphide-bonded loops demonstrate a strongly statistically significant increase in conservation relative to residues immediately adjacent to the cyclic region (Figure 4), however, this increase is very modest in scale, suggesting that disulphide-bonded loop regions may often be embedded within regions of more extended structural constraint. We noted that in our dataset, the increased conservation within disulphide-bonded loops relative to adjacent regions was limited to residues 1,3 and 4 places from the cyclic cysteines and was not seen at the position two residues away. It is not clear why this would be the case, but it is possible that residues with reducing effects may have more influence at certain distances and avoidance of these effects may impact on the pattern of conservation.

It may well be that disulphide-bonded loops with stand-alone function are more likely to be conserved relative to their immediately adjacent residues. Figure 5 shows the distribution of differences in conservation between disulphide-bonded loops and their immediate surrounding residues.

The surface and interface disulphide-bonded loops show a similar distribution to other cyclic regions. Thus, simply existing at a protein surface or interface does not

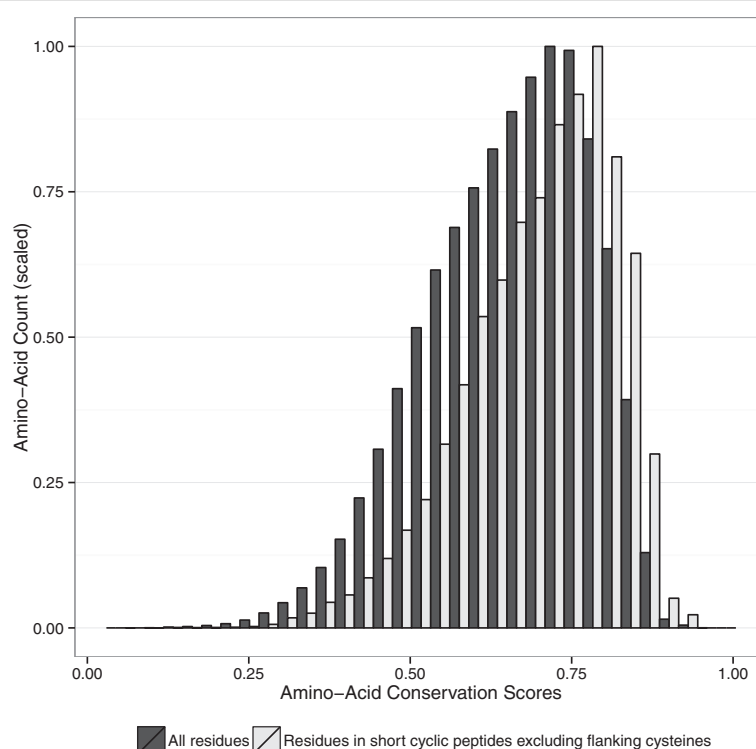


Figure 3 Conservation score distribution of amino acids in proteins containing short disulphide-bonded loops. Amino-acids inside short disulphide-bonded loops are indicated in light grey and the overall distribution is in black. Both distributions are scaled to a maximum Y-axis value of 1. P-value of $< 10^{-15}$ indicates that average conservation inside disulphide-bonded loops is very significantly larger than the overall average conservation, based on a single-tailed Mann-Whitney U test.

mean that a disulphide-bonded loop will be preferentially conserved.

However, the graph does have a smaller second peak at about 0.30, indicating a set of disulphide-bonded loops that are significantly more conserved than their adjacent residues. Table 2 lists the names of the protein families that contain these highly conserved disulphide-bonded loops. Over half are secreted short proteins and peptide hormones such as somatotropin, prolactin, guanylin and urotensin, as well as the larger polygalacturonase from plants. Many of these peptides and proteins are very short, such as guanylin (15 amino acids) and urotensin (11 amino acids), with the disulphide bonded section making up over half of the peptide length. Given the lower conservation outside the cyclic region, such proteins are good candidates for investigating the role of the cyclic regions alone.

Viral proteins as sources of bioactive disulphide-bonded loops

It is known that viruses can rapidly co-evolve many of the same protein-protein regulatory strategies as their hosts [45]. We scanned viral proteins for disulphide-bonded loops that may mimic host disulphide-bonded

loops mediating interactions. Table 4 shows virus proteins containing disulphide-bonded loops, similar to a human disulphide-bonded loop, where there is at least one human-virus protein interaction described in an interaction database.

The match of an adenoviral peptide CNSSTDSC to a human integrin β -6 CTTSTDSC disulphide-bonded loop was of most interest, since viral proteins exploit extracellular matrix and integrin interactions to facilitate cell adhesion and entry. The integrin CTTSTDSC loop is located in the third EGF-like repeat domain region of the extracellular portion of Integrin β -6. This region is known to be “masked” in the integrin’s inactive conformation, and exposed in the active conformation [46]. The human herpesvirus ssDNA binding protein binds the secreted FBLN5 protein: (Additional file 1: Table S2) the viral peptide CRRPC resembles the human ERBB2 CSKPC peptide somewhat. This similarity is relatively weak, but it is intriguing that the FBLN5 protein contains 9 EGF domains, since ERBB2 belongs to a receptor family known to bind EGF domains.

Many of the matches described in Table 4 consist of quite short disulphide loops, 5 or 6 residues long, which points to the possibility of these matches being a chance

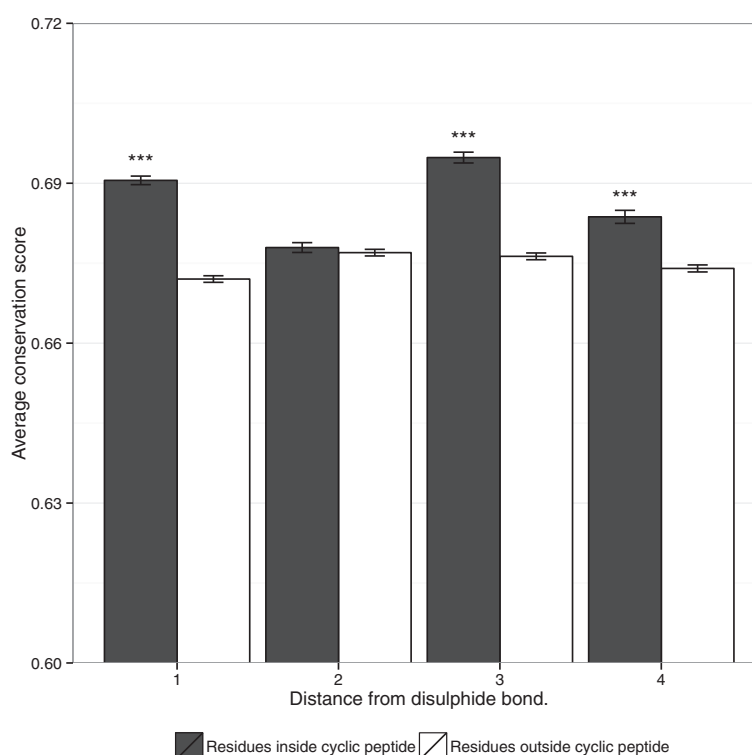


Figure 4 Average amino acid conservation scores close to short loop disulphide bonds. Error-bars indicate standard error. "****" indicates significant results, with p-values < 0.001 based on a Mann-Whitney U-test, where residues at an equal distance from a disulphide bonded cysteine are more conserved inside the disulphide bond.

occurrence. However, as described above, there is some biological support for considering viral proteomes as representing potential sources of cyclic bioactive peptides worthy of further investigation.

Perspectives on disulphide-bonded loops in protein-protein interactions

We would have expected to uncover more disulphide-bonded loops that play key roles in protein-protein interactions, given the importance of cyclic compounds in modulating both enzyme and protein-interactions. It is possible that the selection of small loops within larger proteins is, in general, slower to evolve to the required degree of specificity of interaction than the evolution of larger complementarities between interacting proteins. Thus, in general, proteins do not often seem to evolve structured loops as independent determinants of protein interaction. This is in contrast to eukaryotic Short Linear Motifs (SLiMs) [47] which have specifically evolved to match particular motif binding domains. Why it is that biology has decided to focus on linear motifs rather than cyclic motifs in eukaryotes is not clear. However, it is worth noting that many linear motifs, although they bind independently, typically bind with modest affinity, in a manner that suggests that modest affinity is optimal [48]. Thus,

the benefit of higher affinity binding that cyclic loops might confer may not be advantageous in the context of SLiM signalling, perhaps explaining why our survey has revealed relatively few such interactors. While there are examples of such loops having independent effects from the literature, overall, it may be that mining existing cyclic loop sequences from protein interactions may not be a very fruitful strategy for discovering novel modulators of protein-protein interactions, although we must point out that our own experimental validation only considered one class of such loops.

Conclusions

Considering a small disulphide-bonded loop as a surface-exposed loop pinned in place by a pair of cysteines, the structure of this loop could be relatively independent of the sequence of the larger protein, and mostly determined by the amino acids between the cysteines, due to the conformational constraint imposed by the disulphide bond, and the solvent exposed nature of a surface loop. Thus, a disulphide-bonded loop known to be at a protein-protein interface could on its own, potentially maintain the same binding activity as the parent protein. Since cyclic peptides are known to be more "drug-like" than linear peptides, this approach could provide a promising set

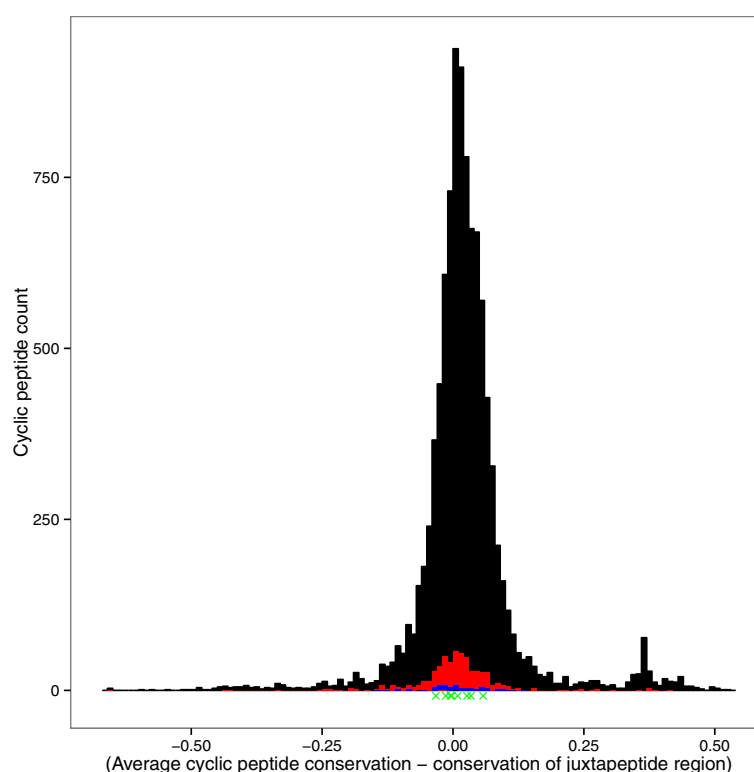


Figure 5 Distribution of the differences between average disulphide-bonded loop conservation scores and the average conservation scores of residues located immediately outside the disulphide-bonded loop. Positive values indicate disulphide-bonded loops more conserved than the regions surrounding them. The black distribution represents all SwissProt annotated disulphide-bonded loops, the red distribution represents disulphide-bonded loops on a PDB protein model surface, and the blue distribution seen at the base of the red distribution represents disulphide-bonded loops at a PDB interface. The green 'X's underneath the histogram represent the conservation difference of the disulphide-bonded loops in Table 1 comprising over 50% of a PDB interface.

of biologically-optimised lead-like molecules to attack the difficult problem of modulating protein-protein interactions [49]. We have contrasted two cases in this study, that of the Bowman-Birk protease derived small macrocyclic peptides, which are known to have independent biological activity, and compared this with a range of EGF-derived small macrocyclic peptides, which do not, showing that while ribosomally derived disulphide-bonded loops can be a promising source of bioactive macrocycles, they may not contain the necessary binding features, even when they make up a majority of a protein-protein interface. Bowman-Birk inhibitors are currently being investigated for their applications in fighting colorectal cancer, due to promising mouse results [50].

A disadvantage of re-purposing disulphide bonded loops as small macrocyclic peptides is that disulphide bonds are more easily broken than the amide bonds present in a traditional head-tail cyclic peptides. Nyugen et al. [51] have compared the serum stability of short (6 residue) antimicrobial linear peptides to the disulphide and head-tail bonded equivalents, and observed that while linear peptides will be 80% degraded in serum over the

course of an hour, disulphide-bonded peptides retain over 50% of their original concentration after 2 hours, and head-tail bonded peptides retain over 70% of their original concentration after 6.5 hours. The concentration of the short depsipeptide arenastatin A in mouse serum has been observed to decline 50% in an hour [52], leading to the conclusion that head-tail bonding is the most desirable from a serum stability point of view. There are a number of well-described strategies which can be used to improve the stability and bioavailability of peptide drugs in the bloodstream [53], and these strategies, along with substituting disulphide bonds for another cyclisation method may be necessary to develop a truly “drug-like” molecule.

Despite being mainly restricted to extracellular or vesicular compartments, short disulphide-bonded loops are relatively widespread with 720 out of a total of 20,252 human proteins in the SwissProt database of manually curated proteins contain annotated disulphide-bonded loops. Disulphide-bonded loop residues are more conserved than non-cyclic residues in the same set of proteins, and are also more conserved than residues located directly beside short disulphide-bonded loops. While the

Table 4 Similar disulphide-bonded loops between human and virus

Human accession	Human peptide	Human protein name	Virus accessions	Virus peptide	Virus protein name
O43506	CELQWC	ADAM20: Disintegrin and metalloproteinase domain-containing protein 20	Q1HVB5	CELGWC	Uncharacterized protein BNLF2b: Human herpesvirus 4
P22413, Q13822	CKGRC	ENPP1 and ENPP2: Ectonucleotide pyrophosphatase/phosphodiesterase family member 1 and 2	O41974	CKRRC	Immediate-early protein, 73: Murid herpesvirus 4
P04626	CSKPC	ERBB2: Receptor tyrosine-protein kinase erbB-2	P03227	CRRPC	Single stranded DNA-binding protein: Human herpesvirus 4
P00451	CRAPC	F8: Coagulation factor VIII	P03227	CRRPC	Single stranded DNA-binding protein: Human herpesvirus 4
P18564	CTTSTDSC	ITGB6: Integrin beta-6	P03254	CNSSTDSC	Early E1A 32 kDa protein: Human adenovirus 2
Q96LB9	CTVSTDC	PGLYRP3: Peptidoglycan recognition protein 3	Q77PU6	CTIRSDC	Protein U90: Human herpesvirus 6B
Q92954, P04004	CKGRC	PRG4: Proteoglycan 4 and VN: Vitronectin	O41974	CKRRC	Immediate-early protein, 73: Murid herpesvirus 4

conservation is modest, there is a subset of peptides which show marked conservation (see Table 2), mainly short secreted peptide hormones or chemical messengers.

This is possibly due to cyclisation by disulphide bonding being the simplest way to impose some structure on a short peptide, which may not have the size to form hydrogen-bonded secondary structural features like helices or beta-sheets. Potential disulphide-bonded loops are also seen in viral proteins, and a subset of these may play potential roles in viral adhesion and entry, or other aspects of viral biology.

This study has used the approach of mining sequence databases for putative disulphide bonded disulphide-bonded loops, conserved relative to their adjacent residues, thus generating a library of compounds of interest. Table 1 contains 13 disulphide bonded loops from well characterised interaction crystal structures, of which only the Bowman-Birk type loops have ever been tested for their independent activity prior to this study. The disulphide bonded loops of interest include not only the loops described in Table 1, but, as the PDB currently only contains structures for a tiny fraction of all possible protein-protein interactions, the 1,231 short disulphide bonded loops at protein surfaces (Additional file 1: Table S3) are also of worthy of investigation, along with the 287 highly conserved disulphide-bonded loops mentioned in Table 2 which are further described in Additional file 1: Table S4.

It is well known that small macrocyclic peptides represent a class of molecules with a wide range of biological activities, and therefore merely showing that there exist

bioactive small macrocyclic peptides derived from larger proteins would not represent any furthering of scientific knowledge. Thus, the key novelty of this work is in exploring how to systematically harness sequence, structural, and evolutionary data over all well characterised proteins in order to identify bioactive proteins. This work both develops a method for identifying potentially bioactive compounds, as well as providing a list of disulphide-bonded loop-protein interaction pairs, readily synthesisable and open to medium or high throughput functional screening for binding or activity.

Methods

Assessing short disulphide-bonded loop structures

Protein data bank (PDB) structures containing short disulphide-bonded loops were identified by searching UNIPROT for manually curated proteins containing a non-overlapping disulphide loop with 2-9 internal residues, with a listed entry in the PDB. The biological assembly format (.pdb1) of each PDB structure was downloaded. Choosing the biological assembly format ensures that the downloaded structure is the biologically relevant form of the structure, as opposed to the crystallographic asymmetric unit, which are not always the same thing. A PyMol [22] script was used to iterate over each structure file, and test each possible pair of cysteine residues spaced up to 11 residues apart in the same protein chain to see if a disulphide bond existed, by checking whether an S-S bond of approximately 2.05 Å existed. For each short disulphide-bonded loop found, the number of surface residues were found by considering

a residue with a solvent accessible surface of over 2.5 Å². The number of residues at a protein-protein interface was found by counting as an interface residue all amino acids with heavy atoms within 3 Å of another protein chain.

EGF receptor activation and competition assays

MCF10A immortalised breast epithelial cells were serum starved for 3 hours before all experiments. For activation experiments, cells were incubated with either 100 μM small macrocyclic peptide or 10 nM EGF for 5 minutes. For EGF/small macrocyclic peptide competition assays, the cells were incubated with 100 μM small macrocyclic peptide for 5 minutes before incubating for another 5 minutes with 1 nM EGF. EGF Receptor activation was then measured by western blotting for phosphorylated Tyrosine 1173 on EGFR, along with blotting for total EGFR and total Actin.

Predicting short disulphide-bonded loops in viruses using protein structural information

Likely short disulphide-bonded loop locations were predicted based on sequence, secondary structure and solvent accessibility information. Secondary structure and solvent accessibility were predicted by the Porter and PaleAle servers from the Distill [54] suite of protein structural prediction servers. Portions of sequences under 11 residues were included that 1) started and ended with a cysteine, 2) contained no internal cysteine residues, 3) were in a region of a protein that was not predicted to be a β sheet or an α helix, and 4) had an average solvent accessibility of the region that was predicted to be more exposed than buried.

Scoring similarity of short disulphide-bonded loops

The similarity of two disulphide-bonded loop sequences was found by aligning the disulphide-bonded loop sequences, excluding the flanking cysteine residues, using the Bio.pairwise2.align function from the BioPython [55] package, which implements pairwise sequence alignment using a dynamic programming algorithm, scored with the BLOSUM62 scoring matrix, and a gap opening and extension penalty of -12. (This gap penalty is three times the BLOSUM62 penalty for the most dissimilar amino-acids, in order to obtain very similar length disulphide-bonded loops). The maximum possible alignment score was calculated by aligning the viral short disulphide-bonded loop to itself, and similarity between human and viral peptides was calculated as

$$\frac{\text{alignment score}}{\text{alignment score}_{\text{max}}}$$

Identifying viral disulphide-bonded loops similar to human proteins

Human and virus proteins containing SwissProt annotated short disulphide-bonded loop regions were identified or predicted using DISTILL as described above. To identify similar viral proteins, we set out to find 1) disulphide-bonded loops with similar sequences in human and viral proteins 2) where the viral proteins had known human interactor proteins 3) where the human proteins interacting with the viral proteins containing a short disulphide-bonded loop also interacted with the original human proteins containing the similar short disulphide-bonded loop.

Uniprot accession numbers were used to uniquely identify proteins, and pairs of Uniprot accessions parsed from the interaction files were used to identify interactions. Similarity between human and virus short disulphide-bonded loop sequences was calculated by aligning and scoring the peptides using the BLOSUM62 substitution matrix to identify similar short disulphide-bonded loops in host and virus. Virus and human short disulphide-bonded loops with a similarity of greater than 0.50 were identified, and checked for shared interactors.

To find which human and virus proteins shared interactors, the sets of known human protein interactions (including human-virus interactions) were downloaded as PSI-MITAB format text files. A text file of human binary protein interactions was downloaded from the MINT [56] database (download date: 08/Apr/2013), which was then parsed to extract human-virus interactions only. A text file of human-virus protein binary interactions was downloaded from the Virhostnet [57] database (download date: 24/Apr/2013), and text files containing human-virus binary interaction data for each of the virus species available in the BioGRID [58] database were also downloaded (download date 24/Apr/2012). All three interaction sources were combined into a single dataset and compared with the list of virus and human proteins with a similarity greater than 0.50, to identify shared interactors.

Additional file

Additional file 1: Supplementary material. Contains supplementary tables and figures referred to in the article.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FC, MD and DS planned the computational survey. FD, DC and DS planned the experiments. FD carried out the computational experiments and wrote the first draft of the manuscript. DC carried out the laboratory experiments. FC, DS, MD and DC revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Science Foundation Ireland, grant number 08 IN.1 B1864 for funding this work.

Author details

¹School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland. ²Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland. ³Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4 Ireland. ⁴Systems Biology Ireland, University College Dublin, Belfield, Dublin 4 Ireland. ⁵Department of Chemistry, Royal College of Surgeons in Ireland, 123 St. Stephens Green, Dublin 2, Ireland.

Received: 3 January 2014 Accepted: 17 July 2014

Published: 17 September 2014

References

- Kotz J: **Bringing macrocycles full circle.** *Science-Business eXchange* 2012, **5**(45). doi:10.1038/scibx.2012.1176
- Gould A, Ji Y, Aboye LT, Camarero AJ: **Cyclotides, a novel ultrastable polypeptide scaffold for drug discovery.** *Curr Pharm Des* 2011, **17**(38):4294–4307. doi:10.2174/138161211798999438
- Lehrer RI, Cole AM, Selsted ME: **θ -Defensins: cyclic peptides with endless potential.** *J Biol Chem* 2012, **287**(32):27014–9. doi:10.1074/jbc.R112.346098
- Hamman JH, Enslin GM, Kotzé AF: **Oral delivery of peptide drugs: barriers and developments.** *BioDrugs: Clin Immunotherapeutics Biopharmaceuticals Gene Ther* 2005, **19**(3):165–77.
- Xiao Q, Pei D: **High-throughput synthesis and screening of cyclic peptide antibiotics.** *J Med Chem* 2007, **50**(13):3132–7. doi:10.1021/jm070282e
- Lamberts SW, van der Lely AJ, de Herder WW, Hofland LJ: **Octreotide.** *N Engl J Med* 1996, **334**(4):246–254. doi:10.1056/NEJM199601253340408
- Schreiber SL, Crabtree GR: **The mechanism of action of cyclosporin A and FK506.** *Immunol Today* 1992, **13**(4):136–142. doi:10.1016/0167-5699(92)90111-J
- Schwarzer D, Finkling R, Marahiel MA: **Nonribosomal peptides?: from genes to products.** *Nat Prod Rep* 2003, **20**(3):275–287. doi:10.1039/b111145k
- Conlan BF, Gillon AD, Craik DJ, Anderson MA: **Circular proteins and mechanisms of cyclization.** *Biopolymers* 2010, **94**(5):573–583. doi:10.1002/bip.21422
- Cascales L, Craik DJ: **Naturally occurring circular proteins: distribution, biosynthesis and evolution.** *Org Biomol Chem* 2010, **8**(22):5035–47. doi:10.1039/c0ob00139b
- Mulvenna JP, Wang C, Craik DJ: **CyBase: a database of cyclic protein sequence and structure.** *Nucleic Acids Res* 2006, **34**(Database issue):192–4. doi:10.1093/nar/gkj005
- Wang CKL, Kaas Q, Chiche L, Craik DJ: **CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering.** *Nucleic Acids Res* 2008, **36**(Database issue):206–10. doi:10.1093/nar/gkm953
- Toniolo C: **Intramolecularly hydrogen-bonded peptide conformations.** *CRC Crit Rev Biochem* 1980, **9**(1):1–44.
- Bisang C, Weber C, Robinson JA: **Protein-loop mimetics: a diketopiperazine-based template to stabilize loop conformations in cyclic peptides containing the NPNA and RGD motifs.** *Helvetica Chimica Acta* 1996, **79**(7):1825–1842. doi:10.1002/hlca.19960790708
- Gunasekaran K, Ramakrishnan C, Balaram P: **Beta-hairpins in proteins revisited: lessons for de novo design.** *Protein Eng Des Select* 1997, **10**(10):1131–1141. doi:10.1093/protein/10.10.1131
- Cochran AG, Tong RT, Starovasnik MA, Park EJ, McDowell RS, Theaker JE, Skelton NJ: **A minimal peptide scaffold for β -turn display: optimizing a strand position in disulfide-cyclized β -hairpins.** *J Am Chem Soc* 2001, **123**(4):625–632. doi:10.1021/ja003369x
- London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O: **Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions?** *Proteins* 2010, **78**(15):3140–9. doi:10.1002/prot.22785
- Xu Y, Rahman NaBD, Othman R, Hu P, Huang M: **Computational identification of self-inhibitory peptides from envelope proteins.** *Proteins* 2012, **80**(9):2154–68. doi:10.1002/prot.24105
- Dechantsreiter MA, Planker E, Mathä B, Lohof E, Hölzemann G, Janczyk A, Goodman SL, Kessler H: **N-Methylated cyclic RGD peptides as highly active and selective α (V) β 3 integrin antagonists.** *J Med Chem* 1999, **42**(16):3033–40. doi:10.1021/jm970832g
- Edwards RJ, Moran N, Devocelle M, Kiernan A, Meade G, Signac W, Foy M, Park SDE, Dunne E, Kenny D, Shields DC: **Bioinformatic discovery of novel bioactive peptides.** *Nat Chem Biol* 2007, **3**(2):108–112.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The protein data bank.** *Nucleic Acids Res* 2000, **28**(1):235–242.
- Schrodinger LLC: **The PyMOL Molecular Graphics System, Version 1.3r1.** 2010. [http://www.pymol.org/citing]
- Cukuroglu E, Gursoy A, Keskin O: **HotRegion: a database of predicted hot spot clusters.** *Nucleic Acids Res* 2012, **40**(Database issue):829–33. doi:10.1093/nar/gkr929
- Nguyen VD, Hatahet F, Salo KEH, Enlund E, Zhang C, Ruddock LW: **Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli.** *Microbial Cell Fact* 2011, **10**:1. doi:10.1186/1475-2859-10-1
- Maupetit J, Derreumaux P, Tufféry P: **A fast method for large-scale de novo peptide and miniprotein structure prediction.** *J Comput Chem* 2010, **31**(4):726–38. doi:10.1002/jcc.21365
- Kihara D, Chen H, Yang YD: **Quality assessment of protein structure models.** *Curr Prot Prot Sci* 2009, **10**(3):216–28.
- Abergel C, Monchois V, Byrne D, Chenivesse S, Lembo F, Lazzaroni J-C, Claverie J-M: **Structure and evolution of the Ily protein family, unexpected lysozyme inhibitors in Gram-negative bacteria.** *Proc Natl Acad Sci U S A* 2007, **104**(15):6394–9. doi:10.1073/pnas.0611019104
- Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, Lee J-O: **Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran.** *Cell* 2007, **130**(5):906–17. doi:10.1016/j.cell.2007.08.002
- Zhou A, Huntington JA, Pannu NS, Carrell RW, Read RJ: **How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration.** *Nat Rev Urol* 2003, **10**(7):541–4. doi:10.1038/nsb943
- Huai Q, Zhou A, Lin L, Mazar AP, Parry GC, Callahan J, Shaw DE, Furie B, Furie BC, Huang M: **Crystal structures of two human vitronectin, urokinase and urokinase receptor complexes.** *Nat Struct Mol Biol* 2008, **15**(4):422–3. doi:10.1038/nsmb.1404
- Alfano D, Franco P, Vocca I, Gambi N, Pisa V, Mancini A, Caputi M, Carriero MV, Iaccarino I, Stoppelli MP: **The urokinase plasminogen activator and its receptor: role in cell growth and apoptosis.** *Thromb Haemost* 2005, **93**(2):205–11. doi:10.1267/THRO05020205
- Sidenius N, Blasi F: **The urokinase plasminogen activator system in cancer: recent advances and implication for prognosis and therapy.** *Cancer Metastasis Rev* 2003, **22**(2-3):205–22. doi:10.1023/A:1023099415940
- Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim J-H, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S: **Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains.** *Cell* 2002, **110**(6):775–787. doi:10.1016/S0092-8674(02)00963-7
- Lockett S, Garcia RS, Barker JJ, Konarev AV, Shewry PR, Clarke AR, Brady RL: **High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds.** *J Mol Biol* 1999, **290**(2):525–33. doi:10.1006/jmbi.1999.2891
- Domingo GJ, Leatherbarrow RJ, Freeman N, Patel S, Weir M: **Synthesis of a mixture of cyclic peptides based on the Bowman-Birk reactive site loop to screen for serine protease inhibitors.** *Int J Pept Protein Res* 1995, **46**(1):79–87. doi:10.1111/j.1399-3011.1995.tb00585.x
- Linggi B, Carpenter G: **ErbB receptors: new insights on mechanisms and biology.** *Trends Cell Biol* 2006, **16**(12):649–56. doi:10.1016/j.tcb.2006.10.008
- Sun YQ: **Glycine residues provide flexibility for enzyme active sites.** *J Biol Chem* 1997, **272**(6):3190–3194. doi:10.1074/jbc.272.6.3190
- Jacob J, Duclouhier H, Cafiso DS: **The role of proline and glycine in determining the backbone flexibility of a channel-forming peptide.** *Biophys J* 1999, **76**(3):1367–76. doi:10.1016/S0006-3495(99)77298-X
- Edwards RJ: **GOPHER: Generation of Orthologous Proteins from High-throughput Evolutionary Relationships.** 2006. [http://bioware.ucd.ie/~compass/biowareweb/]

40. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5):1792–7. doi:10.1093/nar/gkh340
41. Capra JA, Singh M: **Predicting functionally important residues from sequence conservation.** *Bioinformatics (Oxford, England)* 2007, **23**(15):1875–82. doi:10.1093/bioinformatics/btm270
42. Cygler M, Schrag JD, Sussman JL, Harel M, Silman I, Gentry MK, Doctor BP: **Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins.** *Prot Sci: Publication Prot Soc* 1993, **2**(3):366–82. doi:10.1002/pro.5560020309
43. Bustamante CD, Townsend JP, Hartl DL: **Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*.** *Mol Biol Evol* 2000, **17**(2):301–8.
44. Yeh S-W, Liu J-W, Yu S-H, Shih C-H, Hwang J-K, Echave J: **Site-specific structural constraints on protein sequence evolutionary divergence: local packing density versus solvent exposure.** *Mol Biol Evol* 2014, **31**(1):135–9. doi:10.1093/molbev/mst178
45. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR: **Biological and biomedical implications of the co-evolution of pathogens and their hosts.** *Nat Genet* 2002, **32**(4):569–77. doi:10.1038/ng1202-569
46. Shimaoka M, Takagi J, Springer Ta: **Conformational regulation of integrin structure and function.** *Annu Rev Biophys Biomol Struct* 2002, **31**:485–516. doi:10.1146/annurev.biophys.31.101101.140922
47. Gould CM, Diella F, Via A, Puntervoll P, Gemund C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C, Seiler M, Davey NE, Haslam N, Weatheritt RJ, Budd A, Hughes T, Pas J, Rychlewski L, Trave G, Aasland R, Helmer-Citterich M, Linding R, Gibson TJ: **ELM: the status of the 2010 eukaryotic linear motif resource.** *Nucleic Acids Res* 2010, **38** (Database issue):167–80. doi:10.1093/nar/gkp1016
48. Haslam NJ, Shields DC: **Peptide-binding domains: are limp handshakes safest?** *Sci Signal* 2012, **5**(243):40. doi:10.1126/scisignal.2003372
49. Metz A, Pfeleger C, Kopitz H, Pfeiffer-Marek S, Baringhaus K-H, Gohlke H: **Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface.** *J Chem Inf Model* 2012, **52**(1):120–33. doi:10.1021/ci200322s
50. Kennedy AR, Davis JG, Carlton W, Ware JH: **Effects of dietary antioxidant supplementation on the development of malignant lymphoma and other neoplastic lesions in mice exposed to proton or iron-ion radiation.** *Radiat Res* 2008, **169**(6):615–25. doi:10.1667/RR1296.1
51. Nguyen LT, Chau JK, Perry Na, de Boer L, Zaat SAJ, Vogel HJ: **Serum stabilities of short tryptophan- and arginine-rich antimicrobial peptide analogs.** *PloS one* 2010, **5**(9):11–18. doi:10.1371/journal.pone.0012684
52. Murakami N, Tamura S, Wang W, Takagi T, Kobayashi M: **Synthesis of stable analogs in blood and conformational analysis of arenastatin A, a potent cytotoxic spongean depsipeptide.** *Tetrahedron* 2001, **57**(20):4323–4336. doi:10.1016/S0040-4020(01)00339-8
53. Werle M, Bernkop-Schnürch A: **Strategies to improve plasma half life time of peptide and protein drugs.** *Amino Acids* 2006, **30**(4):351–67. doi:10.1007/s00726-005-0289-3
54. Baú D, Martin AJM, Mooney C, Vullo A, Walsh I, Pollastri G: **Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins.** *BMC Bioinformatics* 2006, **7**:402. doi:10.1186/1471-2105-7-402
55. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ: **Biopython: freely available Python tools for computational molecular biology and bioinformatics.** *Bioinformatics* 2009, **25**(11):1422–1423. doi:10.1093/bioinformatics/btp163
56. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G: **MINT: the Molecular INTERaction database.** *Nucleic Acids Res* 2007, **35**(Database issue):572–4. doi:10.1093/nar/gkl950
57. Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, André P, Lotteau V, Rabourdin-Combe C: **VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks.** *Nucleic Acids Res* 2009, **37**(Database issue):661–8. doi:10.1093/nar/gkn794
58. Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M: **The BioGRID interaction database 2013 update.** *Nucleic Acids Res* 2013, **41**(Database issue):816–23. doi:10.1093/nar/gks1158

doi:10.1186/1471-2105-15-305

Cite this article as: Duffy et al.: Computational survey of peptides derived from disulphide-bonded protein loops that may serve as mediators of protein-protein interactions. *BMC Bioinformatics* 2014 **15**:305.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

