ELSEVIER

feature

Drug leads for interactive protein targets with unknown structure

Ariel Fernández[1,2], ariel@afinnovation.com **and L. Ridgway Scott**[3,4]

The disruption of protein–protein interfaces (PPIs) remains a challenge in drug discovery. The problem becomes daunting when the structure of the target protein is unknown and is even further complicated when the interface is susceptible to disruptive phosphorylation. Based solely on protein sequence and information about phosphorylation-susceptible sites within the PPI, a new technology has been developed to identify drug leads to inhibit protein associations. Here we reveal this technology and contrast it with current structure-based technologies for the generation of drug leads.
The novel technology is illustrated by a patented invention to treat heart failure. The success of this technology shows that it is possible to generate drug leads in the absence of target structure.

Introduction

Drug discovery endeavors have been focusing for some time on protein-protein (PP) associations, which are basic molecular events in biology [1]. The recruitment of protein complexes is required to initiate and propagate signaling cascades, regulate enzyme activity, articulate and control mechanistic processes involving molecular motors, and so on. When such associations engage altered binding partners, complex formation can lead to the deregulation of biological functions and drug-based disruption of the aberrant associations could represent new therapeutic opportunities [2–4].

Major problems arise in the identification of drug leads and optimization strategies for small compounds involved in the disruption of PPIs [4]. The latter tend to have low surface curvature and often extend over more than 1000 Å^2 on the protein surface, in contrast to the smaller cavities where natural ligands typically bind [5]. The absence of obvious leads and the sheer size of the binding surface make it difficult to identify candidate compounds that would disrupt PPIs. Despite these obstacles, it is possible in some cases to identify suitable leads and even implement optimization strategies. For example, in the murine double minute 2 (MDM2)/p53 complex, MDM2 has been identified as the E3 ligase responsible for the ubiquitin-related degradation of the master tumor suppressor p53 [6] and the disruption of the MDM2/p53 complex promotes the onset of many cell-fate processes that can halt cancer development and progression [6,7]. In tumors, MDM2 is overexpressed and altered so that cell processes associated with senescence, cell cycle arrest, and apoptosis triggered by p53 activity are suppressed through untiring modulation of p53 via an aberrantly persistent MDM2/p53 complex. In this case, crystallographic structure-based analysis coupled with high-throughput screening has generated useful leads, the so-called 'Nutlins' [7]. These leads ultimately steered the discovery of low-molecular-weight compounds that hold great promise as anticancer agents through the disruption of the MDM2/p53 PPI [6,8].

To address many of the problems related to the epitope size in PP associations, methodologies have been implemented for the identification of 'hot spots' or sites that make a significant contribution to binding [9]. Such approaches are typically based on alanine scanning, assessing the impact of single-residue alanine substitutions (beta-carbon truncation, except for glycines) on the binding free energy. Thus, an effective epitope is determined that is significantly smaller than the PPI and comprises the residues with the most significant contribution to binding [4,10]. Once the size of the epitope has been significantly reduced, fragment-based lead discovery can be utilized to generate promising candidates for competitive binding

1359-6446/© 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.drudis.2015.10.006

[4,11,12]. Once in the optimization phase, the discovery process is aided by biophysical methods for structure analysis, including crystallography, disulfide tethering, surface plasmon resonance, and nuclear magnetic resonance [12].

The problem of therapeutic disruption of a PP association becomes especially difficult when the structure of the targeted binding partner is unknown, because the biophysical methods mentioned above cannot provide useful information. The discovery process becomes even further complicated when the interface can be naturally disrupted by phosphorylation at a specific location on the epitope of one of the binding partners, as is often the case in the regulation of activities that recruit complexes. Even in such cases, we uphold the opinion that it is possible to implement a drug discovery platform based solely on sequence-based predictors of binding epitopes endowed with chemical functionality.

The novel approach hinges on three conceptual tenets that are described in the subsequent sections and contrasted vis-à-vis current methodologies: (i) structural defects in proteins, known as 'dehydrons', promote water exclusion at the interface and, thus, residues paired by dehydrons constitute hot spots promoting protein associations [13]; (ii) dehydrons are identified as order–disorder twilight regions along the protein sequence [14] and, therefore, can be inferred utilizing a sequence-based predictor of intrinsic disorder [15]; and (iii) dehydrons functionalize PPI-disruptive phosphorylation sites in their proximity [16,17]. The efficacy of the technology is subsequently illustrated by a recently patented invention to treat heart failure through disruption of the myosin–Myosin binding protein C (MyBP-C) interface [18]. MyBP-C is a multidomain myosin-binding protein with unreported structure that is a central regulator of cardiac contractility [19,20]. MyBP-C molecules constitute molecular brakes modulating the displacement of myosin motors, with a brake-release mechanism hinging on site-specific phosphorylation. By sequence-based inference of dehydron-rich regions in MyBP-C, drug leads were identified that could be used to cure heart failure [18]. Here, we describe the technological advances utilized in this invention to enable therapeutic disruption of PP associations in a generic context.

Hot spots and dehydron epitopes

The integrity of a soluble protein is contingent on the ability of its structure to exclude water from backbone amide-carbonyl hydrogen bonds. Water-exposed intramolecular hydrogen bonds (dehydrons) constitute structural defects taking the particular form of wrapping deficiencies, as previously described [21]. 'Wrapping' refers to the extent to which the backbone hydrogen bond is shielded from hydration as it is surrounded by side-chain carbonaceous groups. These defects favor removal of surrounding water as a means to strengthen and stabilize the underlying electrostatic interaction and, thus, are predictably implicated in protein associations. By exogenously contributing to the wrapping of preformed hydrogen bonds, PP associations in effect remove the wrapping defects, thereby stabilizing the structure [21].

Intramolecular hydrogen bonds that are not 'wrapped' by a sufficient number of nonpolar groups in the protein itself can become stabilized and strengthened by the attachment of a ligand (i.e., a potential drug) or a binding partner that further contributes to their dehydration. Ample bioinformatics evidence on the distribution of dehydrons at the interface of protein complexes supports this physical picture [22]. Thus, dehydrons have been identified as decisive factors driving the formation of protein complexes.

Dehydrons can be identified from structural coordinates using available software [23], and a code for dehydron identification written as a PyMol open source is presented in [24]. To describe the extent of backbone shielding from hydration, we introduce a quantifier of hydrogen-bond wrapping, 'ρ', indicating the number of nonpolar groups contained within a 'desolvation domain' around the bond. Insufficiently wrapped bonds become deshielded and constitute dehydrons. As discussed above, this approach requires a structure and a way is needed to identify dehydron locations from sequence alone. One alternative is to predict structural coordinates from sequence [25], but unfortunately too many decoys with significant wrapping variability are often generated.

Dehydron-rich regions in soluble proteins are typical hot spots for protein associations because of their propensity towards further dehydration. A functional perspective reinforces this view, because dehydrons constitute vulnerabilities that need to be 'corrected' to maintain the integrity of the protein structure and its functional competence. Thus, specific residues of the binding partner contribute to the stabilizing dehydration of preformed dehydrons as they penetrate their desolvation domain upon association.

A solvent-centric perspective is even more informative about the role of dehydrons as hot spots driving PP associations. The water molecules partially occluded in the dehydron nanoenvironment are frustrated in their hydrogen-bond coordination and, hence, generate interfacial tension. This tension is in turn released upon PP association as the frustrated water molecules are removed from the epitope surroundings [13]. Thus, the residues pairing or significantly wrapping preformed dehydrons in a binding partner are in fact expected and verified to be hot spots driving the PP association [13,21].

Sequence-based prediction of dehydron epitopes

The structural integrity of a soluble protein is contingent on its capacity to exclude water from backbone amide-carbonyl hydrogen bonds. This implies that proteins with dehydron-rich regions must rely on binding partnerships to maintain their structural integrity [26].

Dehydron-rich regions identified on the protein sequence may be characterized as belonging to a 'twilight zone' between order and native disorder [14,22]. This characterization is suggested by a strong correlation between wrapping of intramolecular hydrogen bonds (ρ) and propensity for structural disorder (f_d). The correlation reflects the fact that a local incapacity to exclude water intramolecularly from preformed hydrogen bonds is causative of a local loss of structural integrity, whereby full backbone hydration becomes structurally disruptive.

The local disorder propensity can be accurately quantified by a sequence-based score generated by a predictor of native disorder propensity, such as PONDR-VLXT [15] or other software [27] that takes into account residue attributes, such as hydrophilicity and aromaticity, and their distribution within the window of the protein sequence interrogated [15]. The disorder score ($0 \leq f_d \leq 1$) is assigned to each residue within a sliding window, representing the predicted propensity of the residue to be in a disordered region ($f_d = 1$, certainty of disorder; $f_d = 0$, certainty of order). Only 6% of 1100 nonhomologous Protein Data Bank (PDB) proteins gave false positive predictions of disorder in sequence windows of 40 amino acids [22]. The strong correlation (over 2806 nonredundant nonhomologous PDB domains) between disorder score of a residue and extent of wrapping of the hydrogen bond engaging the residue (if any) implies that dehydrons correspond to structurally vulnerable regions [22]. Hence, the characterization of dehydrons as belonging to the order–disorder twilight range $0.35 < f_d < 0.8$ and flanked by ordered and disordered regions is warranted.

A caveat applies to the use of disorder predictors to infer dehydrons [28]. Dehydrons require detailed information resolved at the residue level, whereas disorder predictors, such as PONDR, provide a smeared out signal as a

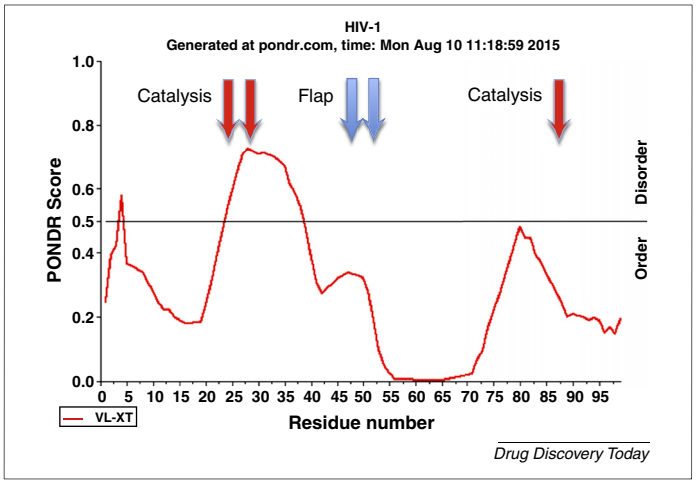

FIGURE 1

Disorder propensity plot for a subunit of the HIV-1 protease functional homodimer generated by PONDR-VLXT [15]. The residues in the order–disorder twilight region representing dehydrons are marked by arrows and correspond to functionally relevant locations on the structure of the homodimeric complex, as shown in Fig. 2.

window of fixed length slides along the protein sequence. In practice, learning technologies for resolution enhancement (LREs), such as Twilighter [28], are required to deconvolute the PONDR signal. The LREs use a training set constructed from protein sequence windows combined with reliable structure-based dehydron identification within the windows and PONDR plots for PDB-reported proteins.

Both order and native disorder are well-characterized structural attributes. However, the vulnerable regions in a soluble structure that promote protein associations [13,14] belong to the novel category of 'tamed disorder'. Thus, neither order nor disorder is an adequate category to describe dehydron-rich regions.

To illustrate the efficacy of sequence-based dehydron inference [15,22] (Fig. 1), we contrast it with structure-based determination, considering the HIV-1 protease in its catalytically competent dimer (PDB 1A30). The free monomers have a large structural twilight content (54%, Fig. 1), as expected given that the protein is natively dimeric. Two dehydron-rich regions are detectable with the software described in [23]: a catalytic region and a flap region (Fig. 2). Dehydrons Asp25–Ala28 ($\rho = 22$) and Ala28–Arg87 ($\rho = 15$) are spatially adjacent to the catalytically competent Asp25, and water exclusion promoted by these dehydrons serves the dual purpose of fostering the association of the peptide substrate and enabling the nucleophilic attack by

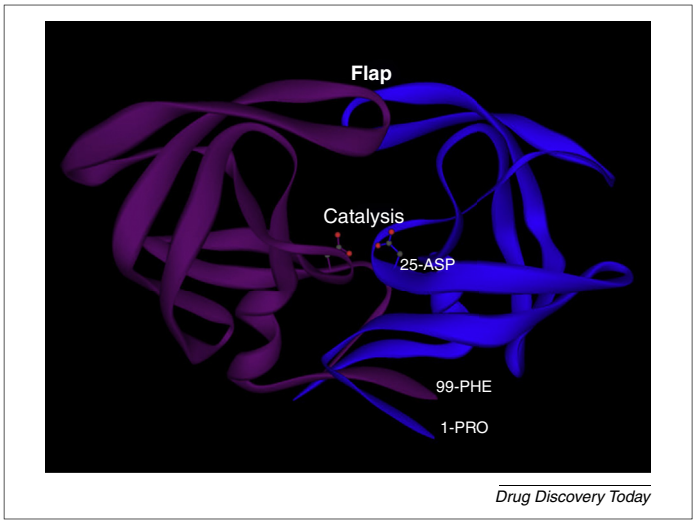

FIGURE 2

Ribbon representation of the functionally competent homodimeric HIV-1 protease (Protein Data Bank 1A30). The individual protein chains are displayed in magenta and blue with catalytic and flap regions indicated. The catalytic residue Asp25 is displayed with a full side chain.

Asp25 by lowering the local dielectric permittivity and preventing the hydrolytic reversal of the enzyme-substrate adduct back to reactants. The flap dehydron Gly49-Gly52 ($\rho = 11$) is required to confer flexibility to the flap region, as required for the mechanistic processivity of the enzyme. In a sequence-based analysis, the dehydrons that shape the catalytic and flap regions can be directly inferred from the order-disorder twilight regions in the PONDR plot for the protease primary sequence (Fig. 1). The flap and catalytic dehydrons lie precisely at the lower and upper boundaries of the twilight region, a common occurrence because PONDR only gives a smeared out signal [28].

Although the sequence-based determination of dehydron-rich regions is less precise than the structure-based approach, it is precise enough to identify lead candidates of interest. As discussed above, a significant improvement of the dehydron signal can be achieved using the Twilighter LRE, as described in [28].

Dehydron epitopes as effectors of phosphorylation-susceptible residues

Dehydrons in the proximity of residues susceptible of becoming phosphorylated have been found to chemically functionalize such residues, enabling their nucleophilic attack on the terminal phosphoester linkage of ATP [16,17]. This chemical role is mediated by an interfacial water molecule around a dehydron that becomes frustrated because of impaired hydrogen bond coordination, thus behaving as a chemical base. Thus, the dehydron-associated interfacial water molecule promotes proton acceptance [17], thereby enhancing the nucleophilicity of a phosphorylation-prone residue in the dehydron proximity. These findings suggest that we can target a PPI susceptible of disruptive phosphorylation by exogenously wrapping the preformed dehydrons that are part of the binding epitope and enable phosphorylation of one of the binding partners.

Drug leads disrupting a PPI in the absence of target structure

As discussed above, dehydrons are main determinants of epitopes, can be inferred from sequence, and can activate residues susceptible of phosphorylation. These properties make it intuitively appealing to identify dehydron-rich regions with peptide-based leads to disrupt PPIs when the structure of the binding partner is unknown and the PPI may be naturally disrupted through site-specific phosphorylation. To illustrate the power of the discovery technology, we focus on the lead identification for disruption of

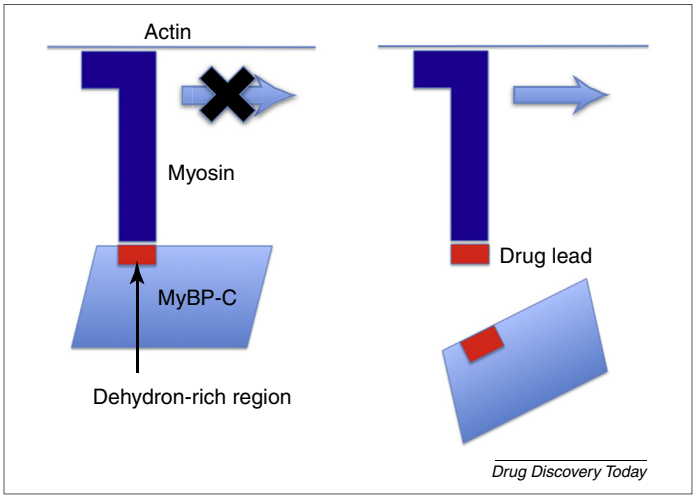

FIGURE 3

Schematic view of the therapeutic disruption of the Myosin binding protein C (MyBP-C)/myosin interface to treat heart failure. The drug lead (red) comprises a small peptide mimicking the dehydron-rich region in MyBP-C that binds to myosin. To prevent detachment, the peptide is modified through Ser302Ala and Ser307Ala substitutions to become nonsusceptible to phosphorylation. As the peptide competitively binds to myosin, it displaces the myosin modulator MyBP-C, enabling the myosin motor to slide along the actin filament, as needed for myocyte contractility. At the molecular level, the therapeutic effect of the drug lead represents an induced 'brake release' for the compromised heart.

a specific PPI, a patented invention en route to cure heart failure [18].

MyBP-C is a central regulator of cardiac contraction [19,20]. In murine models, genetic ablation or phosphorylation of MyBP-C by protein kinase A (PKA) or Calcium/calmodulin-dependent protein kinase II (CAMKII) accelerates contraction and increases its force in cardiac muscle. Most recently, it has been shown that CAMKII phosphorylation of MyBP-C at Ser282 and Ser302 in mice and Ser284 and Ser304 in humans underlies the increase in myocardial contractile force as heart rate is increased, according to the so-called 'staircase phenomenon'. Proof for this statement was obtained by observing that: (i) phosphorylation at these two residues is increased when stimulus frequency is increased, but other potential phosphorylation sites are not; and (ii) the staircase phenomenon is absent in hearts in which these residues are replaced by residues not susceptible to phosphorylation. MyBP-C normally reduces the speed and strength of contraction by means of its interaction with the contractile protein motor myosin and thereby reduces the probability of myosin sliding along actin filaments [19] (Fig. 3). Ablation of MyBP-C or phosphorylation of MyBP-C by PKA or CAMKII disrupts this interaction and relieves MyBP-C repression of myosin [20]. Once phosphorylated, MyBP-C no longer binds to myosin, myosin slides along actin, the probability of myosin binding to actin increases, and the contraction speed and strength increase.

In heart failure, MyBP-C is phosphorylated minimally or not at all because of downregulation of β-adrenergic receptors. Given that phosphorylation of MyBP-C improves contraction, US patent 9,051,387 [18] proposed to target the CAMKII site on MyBP-C with a pharmaceutical designed to disrupt its interaction with myosin and thereby improve cardiac function. This involves designing an optimal MyBP-C-derived peptide. The peptide contains the motif responsible for interaction with myosin and, hence, disrupts the MyBP-C–myosin association as required for therapeutic action (Fig. 3). The myosin-binding peptide-based therapeutic agent can be identified based on the premise that disruption of the myosin–MyBP-C interface would release a molecular brake on cardiomyocyte contractility imposed by the repressive activity of the unphosphorylated form of MyBP-C on myosin.

The 3D structure of the enormous 11-domain protein MyBP-C is unknown and, therefore, it was decided to design the peptide based on the output from PONDR® or another predictor of native disorder [15] (*cf.* Fig. 1). More specifically, the peptide was developed based on a sequence-based prediction of the dehydron-rich region that constitutes the putative myosin-binding site. To predict the peptide sequence, the inventors examined a region between the C1 and C2 domains of MyBP-C in the twilight zone between order and disorder. In this way, it became possible to identify the twilight region containing phosphorylation sites Ser302, Ser307 in the motif region intercalated between domains C1 and C2 of cMyBP-C, and to determine the sequence of the patented peptide ($_{293}$FSSLLKKRDSFRRDSKLF$_{310}$) that could be used as a lead to the therapeutic agent to treat heart failure. The invention has been shown to significantly increase cardiac contractile force and frequency in the failing heart [18]. To turn the peptide into a therapeutic agent still entails an arduous optimization process to improve bioavailability while avoiding peptidase digestion. Optimization strategies would need to be implemented to replace peptide bonds and create unnatural side chains to mimic the key binding components that might emerge from structural characterization of the peptide–myosin association.

Concluding remarks

The drug-based disruption of protein-protein associations involving large multidomain binding partners with unknown structure poses a formidable challenge. Yet, new developments in the sequence-based prediction of protein regions that functionalize interfacial water and create water–protein interfacial tension encourage us to uphold the opinion that it is possible to identify leads to disrupt PPIs even in the absence of a 3D structure.

References

1 Pelay-Gimeno, M. *et al.* (2015) Structure-based design of inhibitors of protein–protein interactions: mimicking peptide binding epitopes. *Angew. Chem. Int. Ed. Engl.* 54, 8896–8927
2 Wells, J.A. and McClendon, C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* 450, 1001–1009
3 Nero, T.L. *et al.* (2014) Onco-genic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* 14, 248–262
4 Arkin, M.R. *et al.* (2014) Small-molecule inhibitors of protein–protein interactions: progressing toward the reality. *Chem. Biol.* 21, 1102–1114
5 Hwang, H. *et al.* (2010) Protein-protein docking benchmark version 4.0. *Proteins* 78, 3111–3114
6 Khoo, K.H. *et al.* (2014) Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat. Rev. Drug Discov.* 13, 217–236
7 Vassilev, L.T. *et al.* (2004) *In vivo* activation of the p53 pathway by small molecule antagonists of MDM2. *Science* 303, 844–848
8 Fry, D.C. (2012) Small-molecule inhibition of protein–protein interactions: how to mimic a protein partner. *Curr. Pharm. Des.* 18, 4679–4684
9 Kouadio, J.L. *et al.* (2005) Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. *J. Biol. Chem.* 280, 25524–25532
10 King, C. *et al.* (2014) Removing T-cell epitopes with computational protein design. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8577–8582

11 Erlanson, D.A. *et al.* (2004) Tethering fragment-based drug discovery. *Annu. Rev. Biophys. Biomol. Struct.* 33, 199–223
12 Winter, A. *et al.* (2012) Biophysical and computational fragment-based approaches to targeting protein–protein interactions: applications in structure-guided drug discovery. *Q. Rev. Biophys.* 45, 383–426
13 Fernández, A. (2012) Epistructural tension promotes protein associations. *Phys. Rev. Lett.* 108, 188102
14 Fernández Stigliano, A. (2015) Proteins in the order-disorder twilight: unstable interfaces promote protein aggregation. In *Biomolecular Interfaces: Interactions, Functions and Drug Design* (Fernandez, A., ed.), pp. 97–124, Springer
15 Xue, B. *et al.* (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochem. Biophys. Acta Proteins Proteomics* 1804, 996–1010
16 Fernández, A. (2014) Communication: chemical functionality of interfacial water enveloping nanoscale structural defects in proteins. *J. Chem. Phys.* 140, 221102
17 Fernández, A. (2015) Packing defects functionalize soluble proteins. *FEBS Lett.* 589, 967–973
18 Moss, R. and Fernández, A. Wisconsin Alumni Research Association (WARF). Inhibition of MyBP-C binding to myosin as a treatment for heart failure, US patent 9,051,387B2.
19 Previs, M.J. *et al.* (2012) Molecular mechanics of cardiac myosin-binding protein c in native thick filaments. *Science* 337, 1215–1218
20 Colson, B.A. *et al.* (2007) Radial displacement of myosin cross-bridges in mouse myocardium due to ablation of myosin binding protein-C. *J. Mol. Biol.* 367, 36–41
21 Fernández Stigliano, A. (2015) The aqueous interface of a soluble protein or the birth of epistructural biology. In *Biomolecular Interfaces: Interactions, Functions and Drug Design* (Fernandez, A., ed.), pp. 1–31, Springer
22 Pietrosemoli, N. *et al.* (2007) Dehydration propensity of order–disorder intermediate regions in soluble proteins. *J. Proteome Res.* 6, 3519–3526
23 Fraser, C.M. *et al.* (2011) *WRAPPA: A Screening Tool for Candidate Dehydron Identification.* University of Chicago
24 Fernández Stigliano, A. (2015) Appendix 1: Code for dehydron identification. In *Biomolecular Interfaces: Interactions, Functions and Drug Design* (Fernandez, A., ed.), pp. 335–341, Springer
25 Chen, J. *et al.* (2007) Molecular basis for specificity in the druggable kinome: sequence-based analysis. *Bioinformatics* 23, 563–572
26 Chen, J. *et al.* (2008) Protein structure protection commits gene expression patterns. *Genome Biol.* 9, R107
27 Zeng, Z. *et al.* (2015) Survey of natural language processing techniques in bioinformatics. *Comp. Math. Meth. Med.* Article ID 674296
28 Fernández Stigliano, A. (2015) Proteins in the order–disorder twilight: unstable interfaces promote protein aggregation. In *Biomolecular Interfaces: Interactions, Functions and Drug Design* (Fernandez, A., ed.), pp. 97–126, Springer

Ariel Fernández[1,2,*]
L. Ridgway Scott[3,4]

[1]*Argentine Institute of Mathematics (IAM), National Research Council (CONICET), Buenos Aires 1083, Argentina*
[2]*AF Innovation, Avenida del Libertador 1092, Buenos Aires 1112, Argentina*
[3]*Department of Computer Science, The University of Chicago, Chicago, IL 60637, USA*
[4]*Department of Mathematics, The University of Chicago, Chicago, IL 60637, USA*

**Corresponding author*: