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

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