Protein wrapping: a molecular marker for association, aggregation and drug design

Ariel Fernández* and Alejandro Crespo

Received 26th June 2008
First published as an Advance Article on the web 15th September 2008
DOI: 10.1039/b804150b

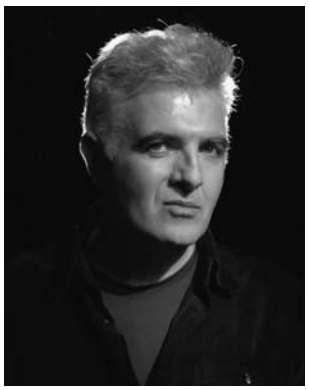
In this *tutorial review* we survey the concept of protein wrapping from a physico-chemical perspective. Wrapping is introduced as an indicator of the packing quality of protein structure. Thus, while a well-wrapped protein is sustainable in isolation, a poorly wrapped protein is reliant on binding partnerships to maintain its structural integrity. At a local level, wrapping is indicative of the extent of solvent exposure of the amide–carbonyl hydrogen bonds of the protein backbone. Poorly wrapped hydrogen bonds, the so-called dehydrons, are shown to represent structural vulnerabilities. These singularities are sticky, hence promoters of protein associations. We also focus on severely under-wrapped protein structures that belong to an order/disorder twilight. Such proteins are shown to be prone to aggregate. Finally, we survey the recent exploitation of dehydrons as targetable features to promote specificity in drug-based cancer therapy. Dehydrons prove to be valuable targets to reduce side effects and enhance drug safety.

1. Introduction

Protein structure in solution is assumed to be sustained by intramolecular forces that are electrostatic and hydrophobic in nature.[1–4] Such forces are actually modulated by an important factor often neglected: the shaping of the solvent microenvironment where they become operational.[5–7] In this regard, we recall that an electrostatic interaction occurring in bulk water is 78 times weaker than the same interaction in an anhydrous medium.[7,8] Thus, the stability and strength of pairwise interactions between different parts of the peptide chain are determined not only by the atomic groups directly engaged in the interaction, but also by the groups involved in shaping their microenvironment. Protecting electrostatic interactions, especially intramolecular hydrogen bonds, from water attack by ''wrapping'' them is just as important as the interactions themselves. The wrapping of an interaction constitutes a way to assess its extent of dehydration. This dehydration is in turn achieved by a clustering of nonpolar groups that exclude water molecules from the surroundings. The relevance of the wrapping concept becomes apparent since the microenvironment of a hydrogen bond determines either the persistence or the ephemeral nature of the interaction and, ultimately, the integrity of the protein structure.[8] In fact, the need for low-dielectric microenvironments for a self-interacting polypeptide chain is central to underpin the physical basis of cooperativity since wrapping constitutes a many-body contribution.[9–11]

The backbone of a protein or peptide chain is highly polar, comprising an amide and carbonyl group per residue. This property introduces constraints on the nature of the hydrophobic collapse and on the chain composition of foldable proteins, *i.e.* those capable of sustaining such a collapse:[12,13] foldable proteins must be able to maintain intramolecular hydrogen bonds by shielding them from the competing backbone hydration. In turn, the thermodynamic forces promoting

Department of Bioengineering, Rice University, 6100 Main St., Houston, TX 77005, USA. E-mail: arifer@rice.edu. E-mail: alecrespo@rice.edu; Fax: 1 713 348 3699; Tel: 1 713 348 3681 (A. Fernández); Tel: 1 713 348 5358 (A. Crespo)



Ariel Fernández

Ariel Fernández received his PhD in Chemistry from Yale University. He currently holds the Hasselmann Professorship in Bioengineering at Rice University. Fernández has published over 300 papers, currently focusing on biophysics and molecular therapeutics.

Alejandro Crespo

Alejandro Crespo is a Postdoctoral Research Associate in the Bioengineering Department at Rice University. He received his doctorate in Chemistry from the University of Buenos Aires, Argentina. His current research focus is the rational design of anticancer drugs.

hydrophobic collapse are known to arise from a minimization of unfavourable protein–water interfaces.[14] Such collapse entails the dehydration of backbone amides and carbonyls, an unfavorable process unless these groups are paired through hydrogen bonding.[14] Only a hydrophobic collapse that ensures the formation and protection of backbone hydrogen bonds is likely to be conducive to sustainable folding.[11] In fact, both collapse and hydrogen bonding are interrelated processes, as corroborated by the correlation between nonpolar surface burial and extent of hydrogen bonding across the protein data bank (PDB).[14]

The hydration of amides and carbonyls competes with the formation of the intramolecular hydrogen bonds. Thus, the structural integrity of proteins is compromised by a "deficiently wrapped" backbone.[11,14] The need for protection of intramolecular hydrogen bonds becomes then a determinant of protein composition.[11]

The strength and stability of backbone hydrogen bonds clearly depend on the microenvironment where they occur: the proximity of nonpolar groups to a hydrogen bond enhances the electrostatic interaction by de-screening the partial charges, and stabilizes it by hindering the hydration of the polar groups in the nonbonded state. Thus, to guarantee the integrity of soluble protein structure, most intramolecular hydrogen bonds must be wrapped by nonpolar groups fairly thoroughly, becoming significantly dehydrated.[11,12]

Taken together, the hydration propensity of amides and carbonyls and the dehydration-induced strengthening of their electrostatic association represent two conflictive tendencies, suggesting that there must be a crossover point in the dehydration propensity of a backbone hydrogen bond. If the bond is poorly wrapped by a few nonpolar groups that cluster around it, then hydration of the paired amide and carbonyl is favorable, but as the hydrogen bond becomes better wrapped, the surrounding water loses too many hydrogen bonding partnerships and thus may be favorably removed.[11]

These generic considerations on the structure of soluble proteins prompt us to review the wrapping concept and the physico-chemical properties associated with poorly wrapped hydrogen bonds. Thereafter, we survey the relevance of wrapping in the contexts of structure vulnerability and protein associations and revisit the aggregation behaviour of extremely under-wrapped proteins. Finally, we survey paradigmatic applications of the wrapping concept in the design of therapeutic drugs.

2. The wrapping of protein structure

As indicated previously, the structural integrity of a soluble protein depends on its packing quality, an indicator of the capacity to exclude water from amide–carbonyl hydrogen bonds.[11,12] Water-exposed intramolecular hydrogen bonds in native folds, the so-called *dehydrons*, constitute structural singularities representing wrapping or packing defects that have been recently characterized.[7,12,15–20] In turn, these defects favor the removal of surrounding water as a means to strengthen and stabilize the underlying electrostatic interaction,[21] and thus are implicated in protein associations[15] and macromolecular recognition.[22,23] The strength and stability of

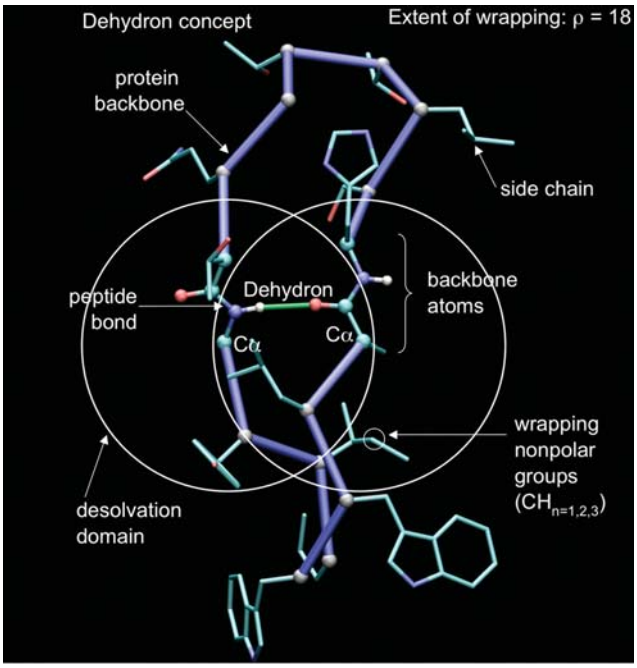


Fig. 1 Dehydron in a soluble protein. The dehydron (ρ = 18), marked in green, pairs two backbone groups (amide and carbonyl, conventional colors for all atoms). The desolvation domain is indicated by two intersecting spheres centred at the α-carbons of the paired residues.

dehydrons may be modulated by an external agent. More precisely, intramolecular hydrogen bonds that are not wrapped by a sufficient number of nonpolar groups may 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 (Fig. 1).[15] Furthermore, dehydrons are established to be decisive factors driving association in 38% of the PDB complexes and constitute significant factors (interfacial dehydron density larger than average on individual partners) in about 95% of all complexes reported in the PDB.[15]

Dehydrons are not the only sticky structural vulnerabilities of soluble proteins: more familiar to structural biologists are the solvent-exposed nonpolar groups on protein surfaces, a destabilizer of the native folded state.[1] Lying outside the folding core, these patches also constitute packing defects and exert a hydrophobic attraction on nonpolar groups of binding partners. As described in the next section, this attraction is comparable in magnitude with the mechanical equivalent (attraction on nonpolar wrappers) of the dehydration propensity of dehydrons.[14–15,21]

Dehydrons may be identified from atomic coordinates of proteins with reported structure, as illustrated in Fig. 1. Thus, we need to introduce an auxiliary quantity, the extent of hydrogen-bond wrapping, ρ, indicating the number of side-chain nonpolar groups (CH_n, n = 1–3) contained within a "desolvation domain" that defines the microenvironment of the bond. The domain is typically defined as two intersecting balls of fixed radius (~thickness of three water layers) centered at the α-carbons of the residues paired by the amide–carbonyl hydrogen bond. In structures of soluble proteins at

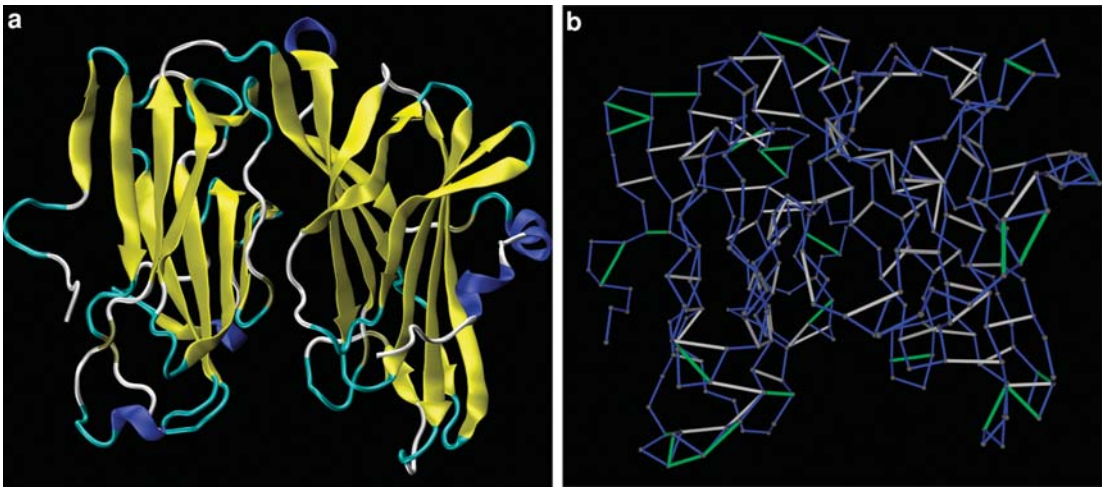


Fig. 2 Illustration of the under-wrapping of protein structure. (a) Ribbon display of the human enzyme peptidylglycine α-hydroxylating monooxygenase (PDB accession code 1SDW). (b) Dehydron pattern for the enzyme protein. Dehydrons are indicated as green segments joining the α-carbons of the paired units, well-wrapped hydrogen bonds ($\rho > 19$) are shown in light grey, and the protein backbone is conventionally shown as blue virtual bonds joining the α-carbons of consecutive amino acid units. The displayed structure has 164 backbone hydrogen bonds, out of which 24 are dehydrons. Thus, the extent of under-wrapping for this protein is 24/164 = 15%.

least two thirds of the backbone hydrogen bonds are wrapped on average by $\rho = 26.6 \pm 7.5$ nonpolar groups for a desolvation ball radius 6 Å. Dehydrons lie in the tails of the distribution, *i.e.* their microenvironment contains 19 or fewer nonpolar groups, so their ρ value is below the mean ($\rho = 26.6$) minus one standard deviation ($\sigma = 7.5$).[16–20]

Thus, the overall under-wrapping of a protein may be assessed by determining the percentage of intramolecular hydrogen bonds with $\rho \leq 19$, that is, the percentage of dehydrons in the structure. An example of the under-wrapping of a protein is given in Fig. 2.

Under-wrapped or dehydron-rich regions in soluble proteins are molecular markers for protein associations because of their propensity for further dehydration.[15] Thus, specific residues of the binding partner containing side-chain nonpolar groups contribute to the desolvation of dehydrons, as they enter the desolvation domain of intramolecular hydrogen bonds upon association. This intermolecular wrapping is illustrated in Fig. 3, displaying the functionally competent human HIV-1 protease dimer. The interface region of the dimer contains 6 dehydrons in the separate monomeric domain. Upon association, specific residues of the binding partner contribute to the desolvation of some of these dehydrons, as they enter their desolvation domains. This intermolecular wrapping reduces the vulnerability of the protease, which has 2 dehydrons in the protein–protein interface upon complexation, instead of the original 6 in the monomeric form (Fig. 3).

3. Dehydron: a sticky wrapping defect

As indicated above, dehydrons have unique physico-chemical properties: they represent structural vulnerabilities since their further hydration disrupts the native conformation, and constitute sticky spots since they promote the removal of surrounding water.[7,12,15,21] How can we demonstrate that a dehydron is likely to attract nonpolar groups in accord with its purported propensity to promote its further dehydration? Previously reported research addressed this question by measuring the adsorption of proteins with wrapping defects onto a hydrophobic surface that may act as a supramolecular wrapper.[21] The experiments entail an evanescent-field spectroscopic interrogation of the hydrophobic surface (a lipid monolayer) enabling a direct measurement of the protein adsorption uptake by detecting changes in the refractive index of the monolayer.[16,21] The monolayer is mounted on a waveguide and interrogated by a light beam of fixed wavelength selected for total reflection along the monolayer–water interface. The refractive changes due to protein penetration are captured by the measured loss of photons arising from refractive distortions of total reflection. The adsorption uptake is shown to correlate tightly with the extent of protein under-wrapping.[21] As an adequate control, only proteins with the same extent of surface hydrophobicity were included. Hence, the attractive drag exerted by dehydrons on test hydrophobes became accessible.[21] The net gain in Coulomb energy associated with wrapping a dehydron has been experimentally determined to be $\sim$4 kJ mol^{-1}.[21] The adhesive force exerted by a dehydron on a hydrophobe at 6 Å distance is $\sim$7.8 pN, a magnitude comparable to the hydrophobic attraction between two nonpolar moieties that frame unfavorable interfaces with water.

This study was motivated by the earlier observations that dehydrons play a pivotal role in driving protein associations, as such associations contribute intermolecularly to stabilize pre-formed structure.[15] In consistency with current terminology, the force stemming from the dehydration propensity of the partially wrapped hydrogen bond is termed *dehydronic*. The dehydronic force arises as a nonpolar group approaches a dehydron with a net effect of immobilizing and ultimately removing surrounding water molecules. This displacement lowers the polarizability of the microenvironment which, in

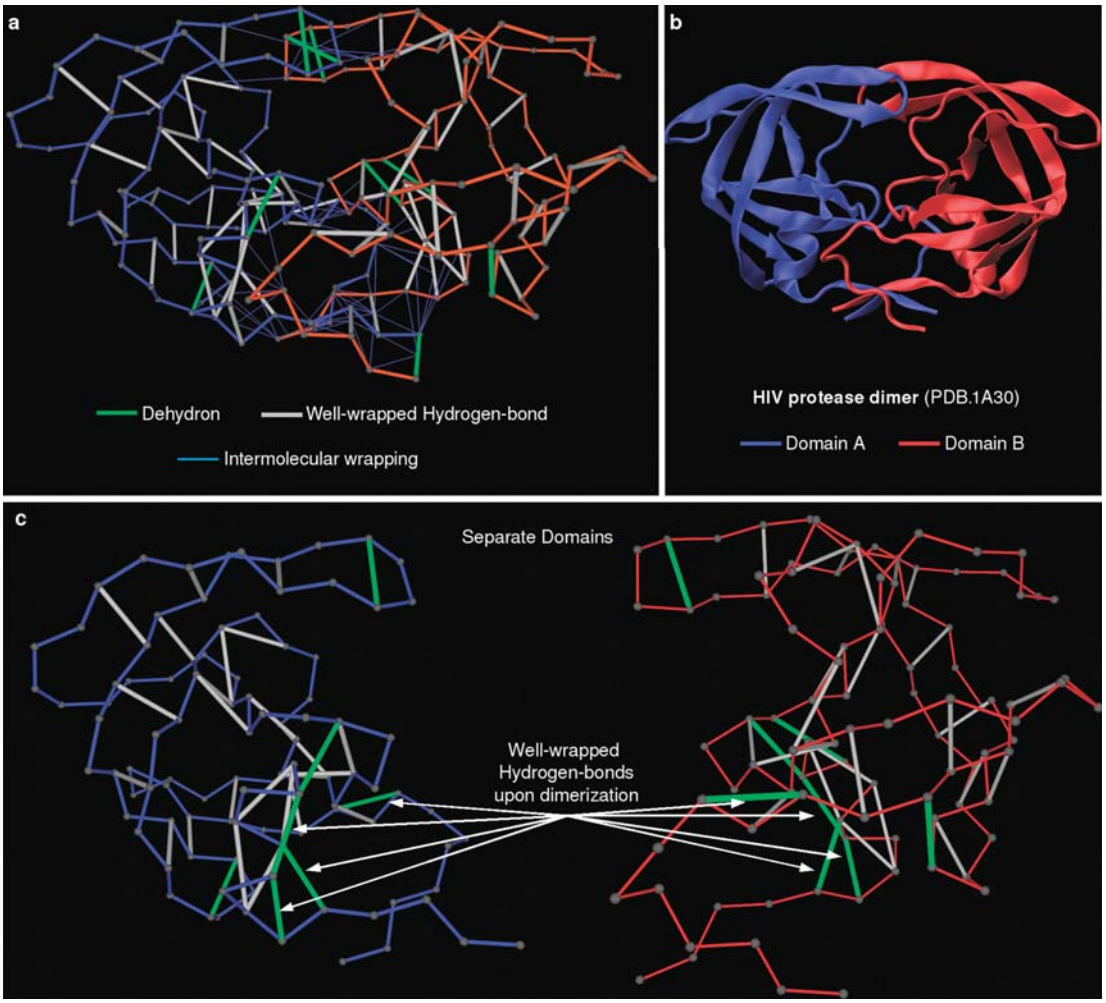


Fig. 3 (a) Intermolecular wrapping in the human HIV-1 protease dimer (PDB.1A30) as a means of protecting the enzyme structure from water attack. Dehydrons are indicated as green segments joining the α-carbons of the paired units, well-wrapped hydrogen bonds are shown in light grey, and the protein backbone is conventionally shown as virtual bonds joining the α-carbons of consecutive amino acid units. Complexation reduces the structural vulnerability of the protease by turning dehydrons into well-wrapped hydrogen bonds through intermolecular protection. This intermolecular wrapping is depicted by thin blue lines from the α-carbon of the wrapping residue to the middle of the hydrogen bond that is protected intermolecularly. (b) Ribbon display of the dimer. (c) Wrapping of separated domains of human HIV-1 protease. Each domain has nine hydrogen-bonds at the interface, six of them being dehydrons. After dimerization, four of these become well-wrapped hydrogen bonds, leaving only two dehydrons at the interface.

turn, de-shields the paired charges.[7,8] Thus, a net attractive force is exerted by the dehydron on a nonpolar group. Since the water molecules solvating an amide and carbonyl paired by a dehydron are necessarily depleted of some hydrogen-bonding partners, the work required for their ultimate removal from the bond surroundings is minimal.[8,21,24] The dehydronic force, denoted $\boldsymbol{\Phi}(\boldsymbol{R})$, is orthogonal to the Coulomb field generated by the polar pair, and may be described within a quasi-continuous treatment by the equation:

$$\boldsymbol{\Phi}(\boldsymbol{R}) = -\boldsymbol{\nabla}_{\boldsymbol{R}}[4\pi\varepsilon(\boldsymbol{R})]^{-1}qq'/r \qquad (1)$$

where $\boldsymbol{R}$ represents the position vector of the hydrophobe or nonpolar group with respect to the center of mass of the hydrogen-bonded polar pair, $\boldsymbol{\nabla}_{\boldsymbol{R}}$ is the gradient taken with respect to this vector, r is the distance between the charges of magnitude q and q' paired by the hydrogen bond (Fig. 4), and the local permittivity coefficient $\varepsilon = \varepsilon(\boldsymbol{R})$ subsumes the

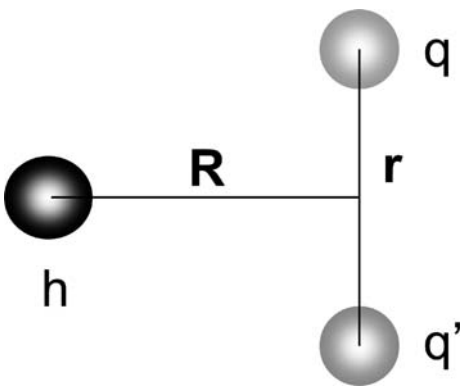


Fig. 4 Orthogonality between dehydronic field exerted on test hydrophobe (h) along coordinate $\boldsymbol{R}$ and the coulomb field exerted along coordinate $\boldsymbol{r}$ between two spherical charges q, q'.

polarizability of the microenvironment, which is generically dependent on the position of the test hydrophobe.[8] An appropriate expression for $\varepsilon(\boldsymbol{R})$ valid at the nanoscales is unavailable at present, because of the discreteness of the dielectric medium and the need to include individual solvent dipole correlations. Although a mean-field dielectric description is unsatisfactory,[8] it is still possible to assert that $\boldsymbol{\Phi}(\boldsymbol{R})$ is an attractive force since a decrease in $||\boldsymbol{R}||$ entails a decrease in local polarization which, in turn, enhances the Coulomb attraction.

4. Extreme under-wrapping, misfolding and aggregation

While wrapping defects in protein structure are recognized markers of protein interactivity,[15,20] a severely under-wrapped protein is prone to misfolding and self-aggregation.[6,8] The misfolding propensity of a severely under-wrapped protein is likely the result of the vulnerability of the native state to water attack. Thus, an excess of 50% dehydrons in the structure of a soluble protein is indicative of a possible misfolder with prion-like functionalities that may lead to aberrant aggregation.[6] The latter often leads to the formation of well-packed and highly periodic supramolecular assemblies, the so-called amyloid fibrils.[25]

The extent of under-wrapping in PDB-reported proteins ranges from 2% to 60%. The cellular form of the human prion PrP^C has 55% dehydrons (Fig. 5). Prions possess an extent of under-wrapping comparable to transmembrane proteins, where under-wrapping does not represent a structural vulnerability as there is no surrounding water. This observation leads to consider the enticing possibility that prions may have been transmembrane proteins displaced to a cytosolic space at some point in their evolution.

C. M. Dobson provided evidence supporting the view that amyloidogenic aggregation, an often pathogenic state of proteins, is actually a generic phase of peptide chains.[26,27] This generic phase is dominated by intermolecular main-chain–main-chain interactions and essentially disregards the information encoded in the primary sequence. Such intermolecular associations appear to be dominated by a basic structural motif: the cross-β structure,[28] an intermolecular sheet-pleated pattern ubiquitous in the fibrous state of aggregation. We should emphasize that this assertion remains conjectural, as no crystal of the fiber for natural prions has been obtained.

While the folding process and its final stable outcome are very much dependent on the amino-acid composition of the chain,[3] the amyloid state appears to be fairly insensitive to the information encoded in the side chains:[25] At first sight, amyloidogenic aggregation does not seem to require an ''aggregation code''. But further analysis reveals that it must place severe constraints on the primary sequence, as some proteins tend to be relatively prone to aggregate even under physiological conditions,[6] while others require extreme conditions to do so or simply do not aggregate reproducibly.[26–29]

Clearly, a selection pressure operates to optimize the primary sequence, so it can render a good folder, *i.e.* an expeditious structure seeker, and a stable soluble structure. This optimization is needed to prevent the functionally competent fold from reverting to a primeval amyloid phase. On the other hand, certain sequences, *i.e.* the ones that yield well-wrapped native structures, are better optimized to escape aggregation than others even under conditions known to sustain the native fold.[6]

While amyloidogenic aggregation has been shown to be always plausible provided sufficiently stringent denaturation conditions are applied,[26–29] a marked amyloidogenic propensity has been detected on a number of proteins under physiological or near-physiological conditions, particularly if the monomeric folding domain is deprived of its natural interacting partners.[16] Such findings imply that not all soluble structures have been optimized to the same degree in order to avert aggregation, and that the more reliant the structure is on binding partnerships or complexations, the more vulnerable it becomes in regards to reverting to the primeval phase. Thus, an over-expression of a folding domain with high complexation requirements, or the modification of its binding partners as a result of genetic accident, or any factor that alters its interactive context are likely to promote the transition to an amyloidogenic state.

These observations inevitably prompt us to address the following question: *to what extent is a severe under-wrapping*

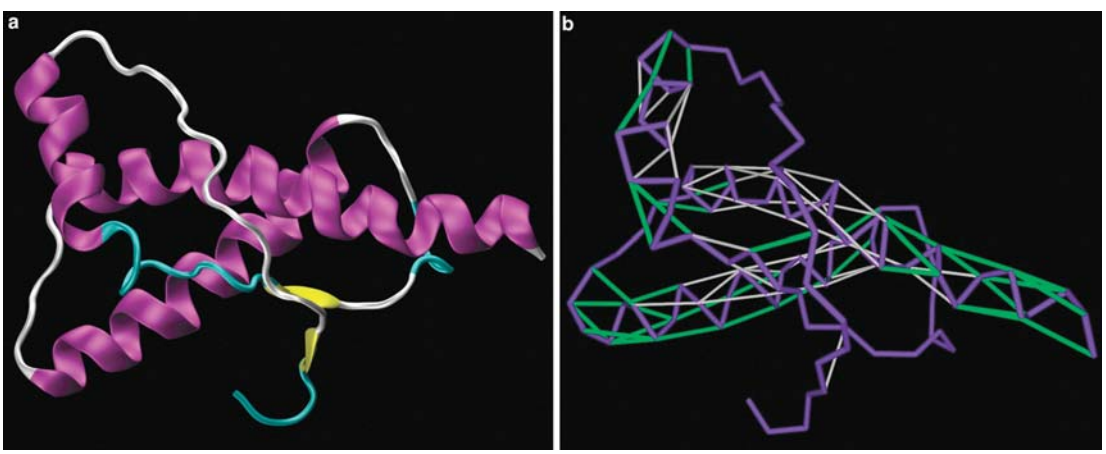


Fig. 5 (a) Ribbon display of the human prion PrP^C in cellular form (PDB.1QM0). (b) Dehydron and hydrogen-bond pattern for the prion protein PrP^C.

of the native fold a good marker for aberrant aggregation? A recent assessment of the *wrapping* of soluble protein structure[6] is critical to address this problem. *For highly under-wrapped proteins (~50% dehydrons), densities higher than 4 dehydrons per 1000 Å^2 of protein surface become inducers of protein aggregation.*[6] This observation turns under-wrapping into a powerful indicator to diagnose amyloidogenic propensity.

Thus, the condition of "keeping the structure dry in water" becomes a requirement to preserve the structural integrity of soluble proteins and imposes a severe building constraint (and thereby an evolutionary pressure) on such proteins. It is expected that the optimization of the structures resulting from this type of evolutionary constraint would be uneven over a range of soluble proteins, resulting in marked differences in aggregation propensity.

The critical role of extreme backbone exposure as an aggregation promoter supports and clarifies the physical picture in which amyloidogenic propensity depends crucially on the dominance of main-chain interactions in detriment of the amino-acid composition that encodes the folded state.[26–29] Precisely, main-chain interactions may dominate if the main chain of the folded state is not properly protected from water attack.[6]

Often, the inability to properly wrap a structure intramolecularly is compensated for by protein complexation. This clarifies the physical picture, suggesting that the more dependent the folding domain is on its interactive partnerships to preserve its structure, the more likely it is to be prone to revert to its primeval aggregated phase.

Direct inspection of the pattern of desolvation of the main chain clearly reveals that the cellular fold of the human prion (Fig. 5) is too vulnerable to water attack and at the same time too sticky to avert aggregation. Clearly, its sequence has not been optimized to "keep the backbone hydrogen bonds dry" in the folded state. In fact, *their extent of exposure of backbone hydrogen bonds is the highest among soluble proteins in the entire PDB*, with the sole exception of some toxins whose stable fold is held together by a profusion of disulfide bridges.[15]

It is suggestive that an inability to protect the main chain is precisely the type of deficiency that best correlates with a propensity to reverse to a primeval aggregation phase determined by main-chain interactions. The actual mechanism by which such defects induce or nucleate the transition is still opaque, although the inherent adhesiveness of packing defects clearly plays a role.

Recently, an atomic-detail structure of a fibrillogenic aggregate, with its β-sheets parallel to the main axis and the strands perpendicular to it, was reported and revealed a tight packing of β-sheets.[30] The cross-β spine of the fibrillogenic peptide GNNQQNY reveals a double parallel β-sheet with tight packing of side chains leading to the full dehydration of intra-sheet backbone–backbone and side-chain–side-chain hydrogen bonds. However, *there is not a single pair-wise interaction between the β-sheets, no hydrogen bonds and no hydrophobic interactions*. Instead, a direct examination of the crystal structure reveals that the association is driven by the dehydration propensity of pre-formed intra-sheet dehydrons, as depicted in Fig. 6.

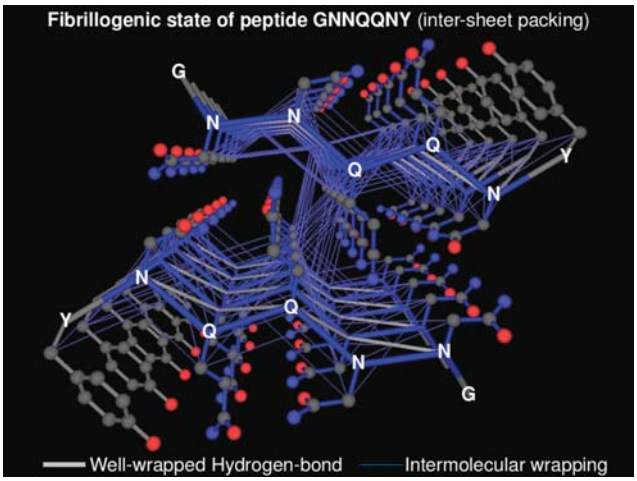


Fig. 6 Inter-sheet wrapping pattern of backbone–backbone hydrogen bonds in the fibrillogenic state of peptide GNNQQNY. The strand backbone is represented as virtual bonds joining α-carbons (blue), and backbone hydrogen bonds are displayed as light grey lines joining the α-carbons of the paired residues. A thin blue line from the β-carbon of a residue to the baricenter of a hydrogen bond indicates wrapping of the bond by the residue: at least one nonpolar group from the residue is contained in the desolvation sphere of the bond.

To conclude, we may introduce some evolutionary considerations. A paradigmatic discovery in biology revealed that folds are conserved across species to perform specific functions. However, the wrapping of such folds is clearly not conserved.[18,19] This fact suggests how complex physiologies may be achieved without dramatically expanding the genome size, a standing problem in biology. Considerable network complexity may be achieved by actually fostering a higher level of complexation or binding partnership, as promoted by a more precarious wrapping of the isolated protein domains. According to our previous analysis, such complex design entails an inherent danger: the reversal of highly under-wrapped folding domains to an amyloidogenic phase even under physiological conditions. Amyloidosis is thus likely to be a consequence of high complexity in proteomic connectivity, as dictated by the structural fragility of highly interactive proteins.[8,20] Thus, the relationships between network centrality, structural wrapping and aggregation propensity probably merit further investigation.

5. Application to molecular targeted therapy

5.1. Drugs as dehydron wrappers

Cancer remains an unsolved purge of modern society. Fundamental concepts are desperately needed. Molecularly targeted drug therapy[31,32] is regarded as one of the most valuable tools in the struggle against malignancy. Thus, cancer-related signals controlling or regulating cell fate and cell proliferation may be blocked through the use of drugs that bind competitively to the signal transducers of the cell: the protein kinases. These kinases possess a natural ligand, ATP (adenine triphosphate), which may be competitively displaced by a man-made drug ligand, thus inhibiting its phosphorylating activity. However, due to the cross reactivity of kinase inhibitors,[33,34]

these treatments typically entail side effects and may become health-threatening. On the other hand, unforeseen cross reactivity has proven to be virtually unavoidable in current combinatorial approaches to drug discovery.[35] Undesirable side effects, arising from drug cross-reactivity and from the diverse roles of the target in different biological scenarios, prompted researchers to advocate for a ''magic-bullet'' paradigm,[36] epitomized by compounds with high binding specificity.

Promiscuous drugs that lack controlled specificity obviously carry a higher risk of inducing life-threatening or severe side effects than their more specific counterparts. If multiple roles of a targeted protein in different cellular contexts are responsible for side effects,[37] it is to be expected that promiscuous compounds would result in an uncertain clinical outcome. Hence, it is forbiddingly dangerous to welcome promiscuous compounds into the therapeutic arena without a rational strategy to control their specificity and therapeutic index. Such control may be achieved if we can identify features in the target structure that are unique to the target, *i.e.* ''selectivity filters'', and drug modifications that promote favorable interactions with such nonconserved features. Thus, selectivity filters may serve as guidance to rational drug design.[38–40]

Kinases, the widespread cancer targets, have paralogs, that is, proteins that share a common ancestor with the target and have diverged away from it after speciation.[41] Thus, kinases belong to common-ancestry groups (families) which typically share the same fold and basic structural features. This structural conservation often results in unexpected cross reactivities that can lead to undesired side effects.[35]

Paralog-discriminating nonconserved features can be exploited as selectivity filters only if they represent targetable differences.[42] These include unique structural features of specific kinases, mostly arising in their inactive conformations.[42] The inactive conformations of kinases make them more discernible, while the active conformation reveals fewer discriminatory features since it is constrained to be catalytically functional. Another strategy to achieve specificity is the design of allosteric kinase inhibitors.[42] These ligands are typically more specific than ATP-competitive inhibitors, since they bind to residues outside the ATP-pocket, which are typically less conserved across paralogs.[42]

A generic selectivity filter of broad applicability arises from comparison of the dehydron patterns of protein targets.[38,41] The dehydron pattern may be turned into an operational selectivity filter for the drug designer for two reasons: (a) dehydrons may be targeted by drugs that further wrap them by bringing nonpolar groups to their proximity upon association (Fig. 1);[38,41,42] and (b) dehydrons are not conserved across paralogs, as illustrated in Fig. 7.[19,41] To assess conservation, we align the paralog structures and compare the microenvironments of the aligned hydrogen bonds. Typically, while

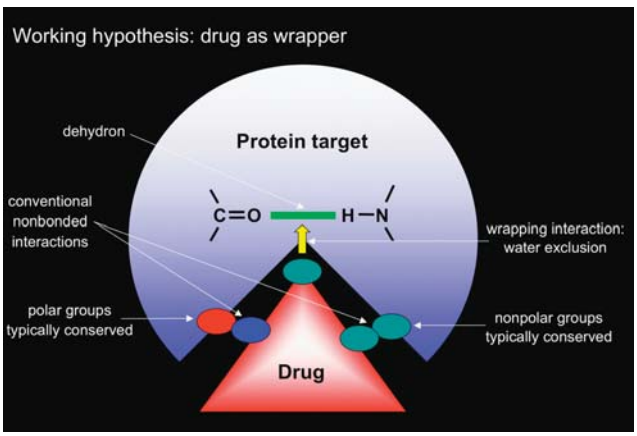


Fig. 8 Cartoon of the design concept of ''drug as wrapper''. Due to the high level of amino acid conservation at the ATP-binding region, intermolecular pairwise interactions (hydrogen bonds, hydrophobic pairing or charge matching) between protein target and drug ligand typically promote promiscuity.[41,42] On the other hand, the exclusion of solvating water from pre-existing solvent-exposed electrostatic interactions (here illustrated with an intramolecular hydrogen bond) may promote specificity if the exposed pre-formed interaction is not conserved across paralogs of the protein target.

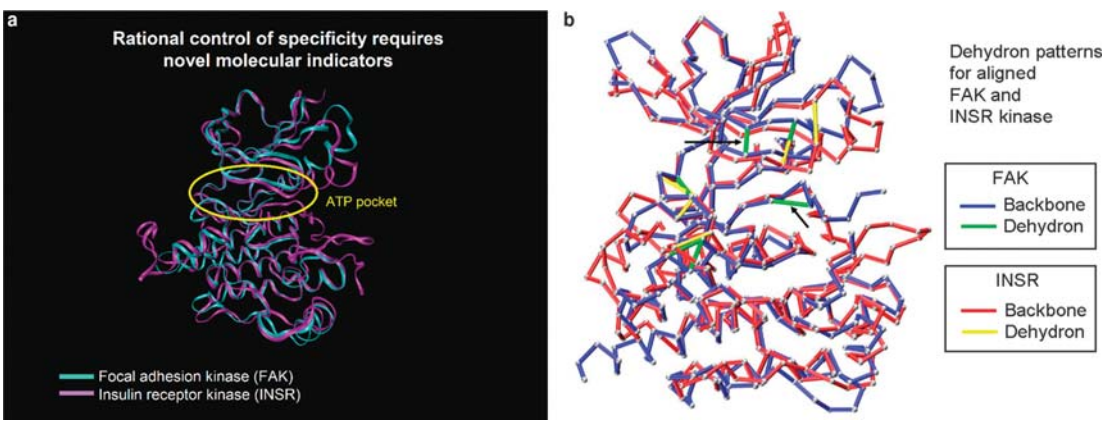


Fig. 7 Wrapping patterns as selectivity filters or discriminators of paralog kinases. This figure illustrates the advantage of exploiting differences in dehydron patterns to discriminate a clinically relevant target implicated in the molecular choreography of cancer metastasis (focal adhesion kinase, FAK) and a target to be avoided (insulin receptor kinase, INSRK). The inhibition of INSRK may have devastating side effects as it impedes glucose uptake and may trigger a diabetic coma. Structural (a) and wrapping (b) alignments of FAK and INSRK. Only the dehydrons on the ATP-pockets are shown. The arrows indicate dehydrons present only in FAK that may targeted in inhibitor design to avoid cross reactivity with INSRK.

the hydrogen bonds are conserved across paralogs, their packing is not, and hence there are differences in the dehydron distribution. Structure conservation across proteins within kinase families enables the alignment.

The wrapping concept introduces a different perspective on drug targets needed to address a problem considered to be the graveyard of most drug-discovery efforts: *the toxicity of side effects*. Side effects may be due to at least two discernible causes: (a) the target protein is involved in different pathways depending on the cellular context, causing the drug/ligand to perturb off-target pathways;[37] (b) the drug/ligand is cross reactive in the sense that its intended target possesses several paralogs which offer potential alternative binding sites because they share essentially the same fold.[41,43] On the other hand, drug discovery remains a semiempirical endeavor, essentially supplemented by structural intuition. Thus, it is unlikely that the high levels of cross reactivity detected in high throughput screening experiments[35] will be tempered or modulated using rational design, unless a new systematic approach is able to discern paralogs above and beyond what a structural characterization may reveal.

Tackling problem (a) directly is too risky and uncertain. It demands a careful assessment of the context dependence of protein connectivity relationships, an area still in its infancy. On the other hand, it makes sense to focus on a strategy to minimize problem (b) by offering a tool of broad applicability to guide drug design based on the nonconserved structure-based indicator of protein interactivity: the dehydron. Dehydrons have been turned into a promising drug-design concept.[38–40] Thus, the concept of ''inhibitor as wrapper of

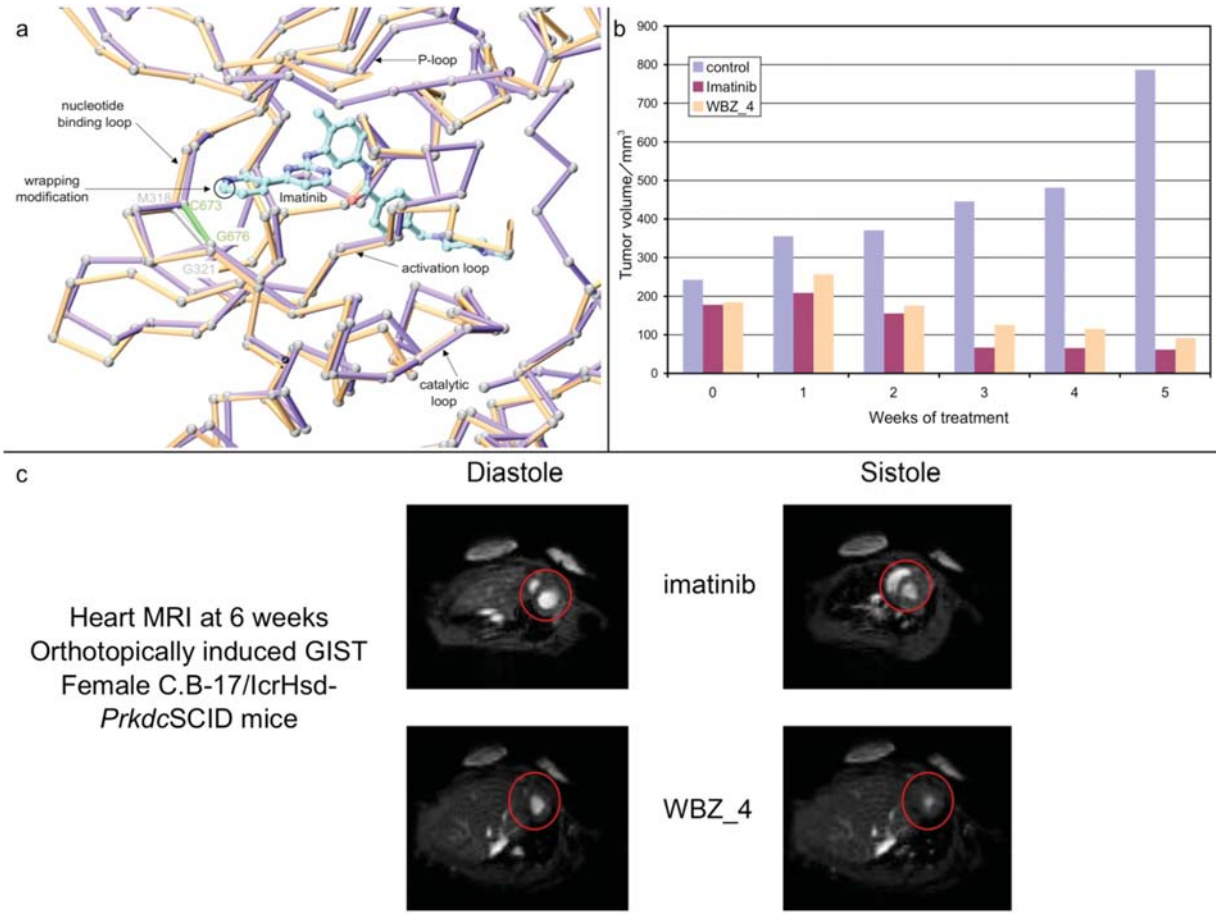


Fig. 9 (a) Drug-guiding wrapping comparison for C-Kit in contact with imatinib (PDB.1T46) aligned with Bcr-Abl kinase (PDB.1FPU). The C-Kit kinase is displayed as violet virtual bonds joining α-carbons aligned with Bcr-Abl kinase (orange virtual bonds). The nonconserved dehydron (green virtual bond joining α-carbons) C673–G676 in C-Kit, aligns with the well-wrapped M318–G321 hydrogen bond (grey virtual bond) in Bcr-Abl, and has been targeted by a methylation wrapping modification of imatinib to achieve specificity. The difference in dehydron patterns for the two primary imatinib targets leads us to develop the re-engineered compound WBZ_4.[38] (b) Anticancer activity of WBZ_4 compared with imatinib on a GIST animal model.[38] Xenograft models of anticancer activity reveal that WBZ_4 is as effective as imatinib therapy on *in vivo* GIST growth determined by longitudinal tumor volume and weight measurements. Mice were randomized to treatment with either control (normal PBS and empty liposomes give indistinguishable results within experimental uncertainty), imatinib or liposome-formulated WBZ_4. WBZ_4 selectivity was corroborated *in vivo* through its lack of antitumor activity on a xenograft of CML K562 tumor cells.[38] (c) Reduced cardiotoxicity of WBZ_4 when compared with imatinib. Comparison of left ventricular cavity in GIST-induced mice treated with WBZ_4 *versus* mice treated with imatinib. Left ventricular ejection fraction (LVEF) was obtained from MRI cross section analysis after 6 weeks of control (groups treated with either PBS or empty liposomes), imatinib, or WBZ_4 therapy in GIST-induced mice. The LVEF for imatinib was 41 ± 8%, while that for WBZ_4 was 59 ± 7%, with control values at 56 ± 8%.[38]

protein packing defects'' (Fig. 8) becomes a pivotal assumption in the bioinformatics tool we propose to develop to guide the design of a new generation of highly selective and safer kinase inhibitors.

5.2. Drug as dehydron wrapper: proof of concept

We have experimentally substantiated the design concept of ''inhibitor as wrapper of packing defects''[38,42,43] and this places us in a position to tackle central therapeutic problems involving selective drug-based inhibition of cancer-related kinases.[35] Enabling technologies will emerge as molecular wrappers of packing defects are turned into efficient selective inhibitors. The paradigm of drug as dehydron wrapper may prevail for at least three reasons: (a) because dehydrons differentiate evolutionarily related proteins representing alternative drug targets; (b) because we are able to target dehydrons with drug ligands that serve as their wrappers; and (c) because the selectivity of commercially available drugs is in fact governed by the dehydron pattern of the kinases, although the drugs were not purposely design to wrap dehydrons.[41]

Here we survey the wrapping technology at work, to demonstrate the feasibility of successfully designing a wrapping inhibitor that is able to discriminate kinase targets based on their differences in dehydron patterns. This problem translates well into a long-standing problem in rational drug design: *the re-engineering of a drug to reduce its side effects when the latter arise from uncontrollable cross reactivity*.

Thus, we shall focus on the powerful anti-cancer drug imatinib (Gleevec, STI571)[44,45] and the control of its side effects[37] through a redesign guided by the wrapping concept. Imatinib's potential cardiotoxicity has been traced to its inhibitory impact in cardiomyocites on its primary target Abelson kinase (Abl).[37] On the other hand, imatinib is known to be effective in treating chronic myeloid leukemia (CML), by inhibiting the deregulated Bcr-Abl kinase, a constitutively active chimera resulting from a chromosomal translocation that fuses the *Bcr* and *Abl* genes.[44,45] Inspired by these observations, imatinib has been re-engineered[38] based on the wrapping concept to curb its potential cardiotoxicity.[37] The re-engineered compound was intended and has been demonstrated at the pre-clinical level to serve as a therapeutic agent to treat gastrointestinal stromal tumor (GIST) through its inhibitory impact focused on the C-Kit kinase, another primary target for imatinib,[46] while its cardiotoxicity is severely reduced when compared with imatinib's effect.[38]

The imatinib redesign, named WBZ_4, was guided by across-target differences in the dehydron patterns of ligand-binding regions.[38] Thus, we sculpted a discriminating modification of imatinib that hampers Bcr-Abl inhibition and re-focuses the impact on C-Kit kinase. We delineated the molecular blueprint for target discrimination through *in vitro* assays and through a phage-display kinase screening library.[35] By assaying for antitumor activity on different cell lines and through immunoblots, we demonstrated controlled inhibitory impact *in vivo*. Finally, we established the therapeutic impact of the engineered compound in a novel GIST animal model and corroborated a significant reduction in cardiotoxicity.[38]

Thus, Fig. 9 shows three milestones in the wrapping-drug development.[38] Fig. 9a shows the difference in dehydron patterns between the two primary imatinib targets, C-Kit and Abl, and the location of the discriminating wrapping modification (a methylation) intended to refocus the inhibitory impact on C-Kit while avoiding impacting Abl. The lack of inhibitory activity against the latter is the result of steric hindrance due to the intramolecularly well packed hydrogen bond in the Abl kinase. Fig. 9b displays the *in vivo* assay for GIST antitumor activity showing comparable therapeutic efficacy for both WBZ_4 and imatinib.[38] Fig. 9c illustrates the cardiotoxicity MRI assay on the GIST-induced animals treated with WBZ_4, adopting imatinib treatment as control.

This example suggests a means of exploiting the wrapping concept to re-design and re-optimize a major clinical lead in order to control its specificity and refocus its inhibitory impact.[38–40]

6. Conclusions

Combining a physico-chemical and structural perspective, we surveyed protein wrapping as a descriptor of packing quality and described a structural singularity associated with the wrapping concept: the dehydron. This singularity represents a vulnerable region in the protein structure and also a sticky spot, hence behaving as a promoter of protein associations. These properties led us to the exploration of two different but interrelated contexts: severely under-wrapped proteins with a tendency to misfold and aggregate, and applications of the wrapping concept in drug design.

We surveyed foundational results that herald a promising technology in drug discovery: *the wrapping technology*. This technology focuses on unique vulnerabilities in the structure of proteins, targeting them with drugs that effectively function as ''dehydron wrappers''. Thus, a design approach is likely to emerge to tackle the long-standing problem of reducing side effects in molecular therapy.

Acknowledgements

The research of A. F. is supported by NIH grant R01-GM072614.

References

1 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1–63.
2 G. Nemethy and H. A. Scheraga, *J. Chem. Phys.*, 1962, **36**, 3382–3401.
3 C. B. Anfinsen, *Science*, 1973, **181**, 223–230.
4 C. Tanford, *Science*, 1978, **200**, 1012–1018.
5 A. Fernández, *J. Chem. Phys.*, 2001, **114**, 2489–2502.
6 A. Fernández, J. Kardos, R. Scott, Y. Goto and R. S. Berry, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6446–6451.
7 A. Fernández, *Nat. Biotechnol.*, 2004, **22**, 1081–1084.
8 N. Pietrosemoli, A. Crespo and A. Fernández, *J. Proteome Res.*, 2007, **6**, 3519–3526.
9 A. Fernández, A. Colubri and R. S. Berry, *Physica A*, 2002, **307**, 235–259.
10 A. Fernández, A. Colubri and R. S. Berry, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 14062–14066.
11 A. Fernández, T. R. Sosnick and A. Colubri, *J. Mol. Biol.*, 2002, **321**, 659–675.
12 A. Fernández and R. Scott, *Biophys. J.*, 2003, **85**, 1914–1928.

13 F. Avbelj and R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5742–5747.
14 A. Fernández, J. Kardos and J. Goto, *FEBS Lett.*, 2003, **536**, 187–192.
15 A. Fernández and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 113–118.
16 A. Fernández and R. S. Berry, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 2391–2396.
17 A. Fernández, *J. Mol. Biol.*, 2004, **337**, 477–483.
18 A. Fernández, R. L. Scott and R. S. Berry, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 2823–2827.
19 A. Fernández and R. S. Berry, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 13460–13465.
20 A. Fernández, *PLoS Comput. Biol.*, 2007, **3**, e226.
21 A. Fernández and L. R. Scott, *Phys. Rev. Lett.*, 2003, **91**, 018102.
22 C. Deremble and R. Lavery, *Curr. Opin. Struct. Biol.*, 2005, **15**, 171–175.
23 B. Ma, T. Elkayam, H. Wolfson and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5772–5777.
24 A. Fernández, *J. Chem. Phys.*, 2003, **119**, 5315–5319.
25 S. B. Prusiner, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 13363–13383.
26 C. M. Dobson, *Trends Biochem. Sci.*, 1999, **24**, 329–332.
27 C. M. Dobson, *Philos. Trans. R. Soc. London, Ser. B*, 2001, **356**, 133–145.
28 M. Sunde and C. C. F. Blake, *Q. Rev. Biophys.*, 1998, **31**, 1–39.
29 C. M. Dobson, *Nature*, 2002, **418**, 729–730.
30 R. Nelson, M. Sawaya, M. Balbirnie, A. Madsen, C. Riekel, R. Grothe and D. Eisenberg, *Nature*, 2005, **435**, 773–778.
31 B. J. Druker, *Oncologist*, 2004, **9**, 357–360.
32 A. Levitski and A. Gazit, *Science*, 1995, **267**, 1782–1788.
33 J. Couzin, *Science*, 2005, **307**, 1711.
34 A. L. Hopkins, J. S. Mason and J. P. Overington, *Curr. Opin. Struct. Biol.*, 2006, **16**, 127–136.
35 M. A. Fabian, W. A. Biggs, D. K. Treiber, C. E. Atteridge, M. D. Azimioara, M. G. Benedetti, T. Carter, P. Ciceri, P. T. Edeen, M. Floyd, J. M. Ford, M. Galvin, J. L. Gerlach, R. M. Grotzfeld, S. Herrgard, D. E. Insko, M. A. Insko, A. Lai, J.-M. Lélias, S. Mehta, Z. V. Milanov, A. M. Velasco, L. M. Wodicka, H. K. Patel, P. P. Zarrinkar and D. A. Lockhart, *Nat. Biotechnol.*, 2005, **23**, 329–336.
36 B. L. Roth, D. J. Sheffler and W. K. Kroeze, *Nat. Rev. Drug Discovery*, 2004, **3**, 353–359.
37 T. Force, D. S. Krause and R. A. Van Etten, *Nat. Rev. Cancer*, 2007, **7**, 332–344.
38 A. Fernández, A. Sanguino, Z. Peng, E. Ozturk, J. Chen, A. Crespo, S. Wulf, A. Shavrin, C. Qin, J. Ma, J. Trent, Y. Lin, H. Han, L. S. Mangala, J. A. Bankson, J. Gelovani, A. Samarel, W. Bornmann, A. K. Sood and G. Lopez-Berestein, *J. Clin. Invest.*, 2007, **117**, 4044–4054.
39 G. Demetri, *J. Clin. Invest.*, 2007, **117**, 3650–3653.
40 S. Crunkhorn, *Nat. Rev. Drug Discovery*, 2008, **7**, 120–121.
41 J. Chen, X. Zhang and A. Fernández, *Bioinformatics*, 2007, **23**, 563–572.
42 A. Crespo and A. Fernández, *Drug Discovery Today*, 2007, **12**, 917–923.
43 A. Fernández, *Structure*, 2005, **13**, 1829–1836.
44 S. Dan, M. Naito and T. Tsuruo, *Cell Death Differ.*, 1998, **5**, 710–715.
45 N. J. Donato and M. Talpaz, *Clin. Cancer Res.*, 2000, **6**, 2965–2966.
46 S. Attoub, C. Rivat, S. Rodrigues, S. Van Borexlaer, M. Bedin, E. Buyneel, C. Louvet, M. Kornprobst, T. Andre, M. Mareel, J. Mester and C. Gespach, *Cancer Res.*, 2002, **62**, 4879–4883.

