Intrinsically Disordered Proteins 1, e26790; 2013; © 2013 Landes Bioscience

Designing disorder
Tales of the unexpected tails

David P Minde[1,*], Els F Halff[2], and Sander J Tans[1]

[1]FOM institute AMOLF; Amsterdam, The Netherlands; [2]Crystal and Structural Chemistry; Bijvoet Center for Biomolecular Research; Utrecht University; Utrecht, The Netherlands

Keywords: fusion tag, disorder prediction, reversible aggregation, protein folding, protein-protein interactions, hydration radius

Protein tags of various sizes and shapes catalyze progress in biosciences. Well-folded tags can serve to solubilize proteins. Small, unfolded, peptide-like tags have become invaluable tools for protein purification as well as protein-protein interaction studies. Intrinsically Disordered Proteins (IDPs), which lack unique 3D structures, received exponentially increasing attention during the last decade. Recently, large ID tags have been developed to solubilize proteins and to engineer the pharmacological properties of protein and peptide pharmaceuticals. Here, we contrast the complementary benefits and applications of both folded and ID tags based on predictions of ID. Less structure often means more function in a shorter tag.

Introduction

IDPs are abundant in all eukaryotic proteomes.[1] Short ID regions mediate some 40% of all protein-protein interactions, for instance in regulatory tails of proteins.[2,3] Large ID regions occur in essential scaffolds and molecular chaperones.[4-7]

What makes IDPs special? IDPs don't form rigid hydrophobic cores due to lower fractions of hydrophobic and aromatic residues (W, I, Y, F, L, H, V).[8] They are often more elongated due to a predominant self-repulsive charge (E or K) or disorder-promoting hydrophilic (S) or structure-breaking residues (P, G).[8,9] Enlarged accessible surfaces of IDPs can facilitate multiple types of interactions including phosphorylation by protein kinases.[10,11] The ~500 human kinases regulate crucial cellular processes including cell cycle and cell differentiation by phosphorylation at thousands of sites, which are typically part of short linear motifs.[12]

Whereas X-ray crystallography has transformed our structural understanding of folded proteins, still relatively little is known about conformational ensembles of IDPs. Unique formidable challenges in IDP research start at the stage of soluble overproduction and subsequent protein purification.[13] High susceptibility to proteolysis or a high aggregation-propensity of many IDPs has slowed down progress in the IDP research field.[14,15] Improvements in solubilizing IDPs, facilitating their purification and efficient prevention of their aggregation, have the potential to catalyze future progress in the IDP research field, which is already growing exponentially.[16]

Protein tags of various sizes and shapes have helped solubilizing, purifying and characterizing proteins (**Table 1**).[17,18] Fusing a folded tag may solubilize proteins during production and purification.[19] Multiple small tags mediate reversible to tight or even covalent binding during protein purification and subsequent characterization and applications of the purified proteins.[20,21] Here we review these commonly used folded and ID tags with an emphasis on contrasting their known structure *vs* known or predicted ID propensity and applications. Finally, we aim to highlight recently emerging applications of large designed ID tags. Large ID tags have the potential to transform protein production and purification, protein-based medicine and to make medicine more patient-specific.[22-25]

Folded Tags Facilitate Production, Purification, and Rapid Labeling In Vivo

Folded fusion tags, ranging in size from ~5 kDa to 70 kDa, are versatile tools that serve to increase folding and solubility during recombinant protein production or to enable fluorescent tracing of single proteins.

Commercially available examples of solubility-enhancing folded protein fusion tags include MBP, NusA and GST.[26,27] These tags significantly improved the soluble overproduction of a range of target proteins on a proteomic scale.[19] Similar effects have been described more recently for the smaller ySUMO, Lipoyl, and Trx fusions, which seem to be of particular use for the overproduction of peptides and small IDPs.[28,29] Fusion of molecular chaperones, chiefly of the Hsp70 class, helped solubilizing aggregation-prone IDPs. For instance, the highly aggregation-prone prion protein could be solubilized during overproduction by fusing it to DnaK, the *E. Coli* homolog of Hsp70.[30]

In addition to their role in protein production, well-folded proteins can mediate a large number of specific functions, including fluorescence. Green fluorescent protein (GFP) and its variants are popular genetically encoded fluorescent labels, which generally do not interfere with the function of their fusion partner.[31] An elegant application of folded tags is the so-called split-FP assay, used in protein-protein interaction studies, where

*Correspondence to: David P Minde; Email: d.minde@amolf.nl
Submitted: 10/10/2013; Revised: 10/11/2013; Accepted: 10/11/2013;
Published Online: 01/01/2013
http://dx.doi.org/10.4161/idp.26790

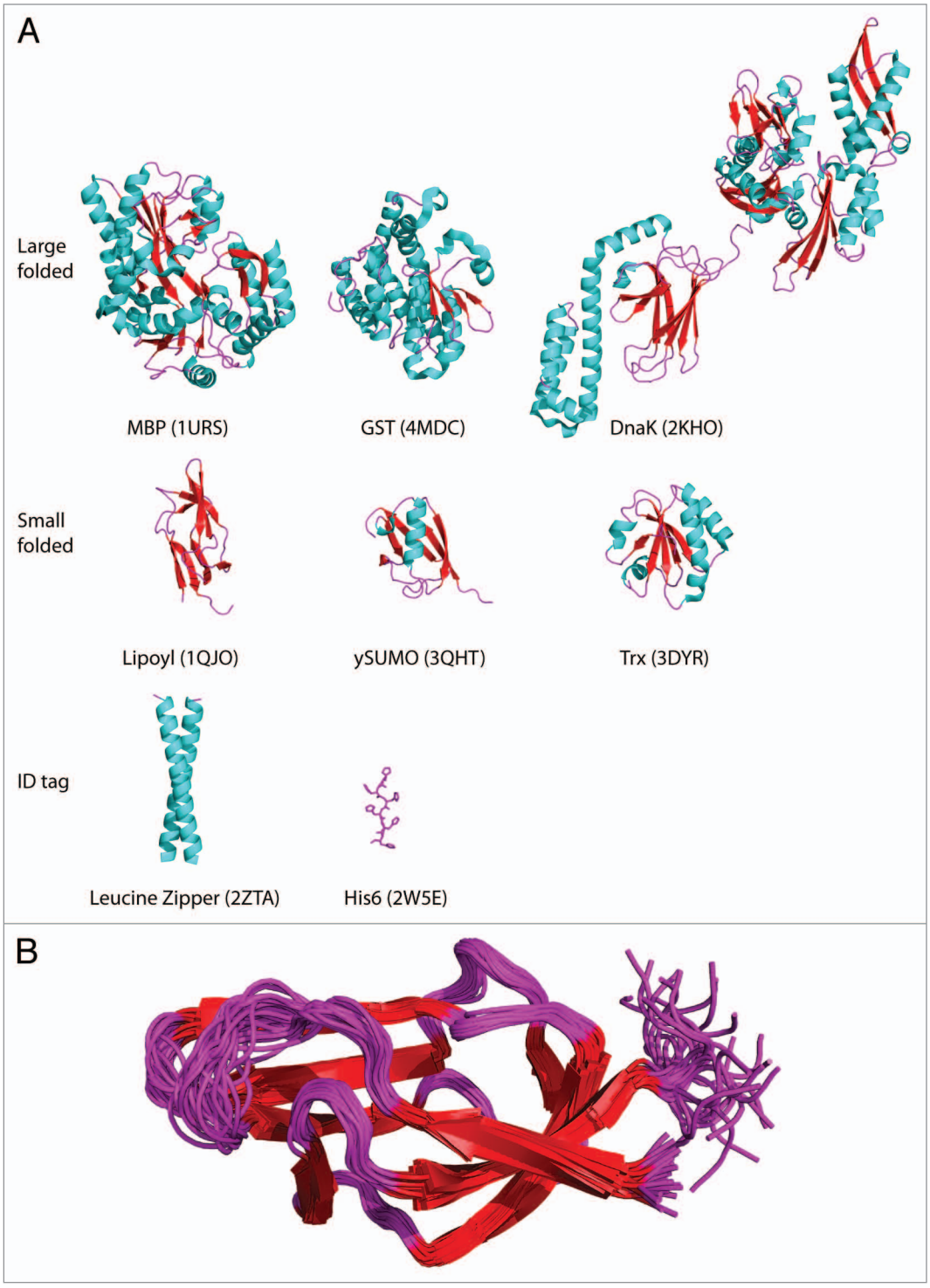

Figure 1. Structural representation of tag order and disorder. (**A**) Cartoon representation of selected protein fusion tags as indicated in the text and **Table 1**. His_6 tag is shown in stick representation. All tags are shown at the same scale. Numbers between brackets indicate the PDB ID from which the represented structures were derived. The fusion tags are colored by secondary structure: cyan indicates an α-helix, red indicates a β-sheet, and purple indicates a loop or turn. (**B**) Overlay of the different conformations of lipoyl NMR structures (PDB ID: 1QJO), showing limited flexibility.

2 halves of GFP are appended to putatively interacting proteins to map their interactions in space and time in living cells. Only upon close encounter of 2 proteins, both fused GFP halves complement each other to form the mature fluorophore. A recent application of split-GFP to trace direct molecular contacts of proteins involved in peroxisome-biogenesis demonstrated that this approach yields super-resolved insights on the low nm scale using a conventional wide-field fluorescence setup.[32,33]

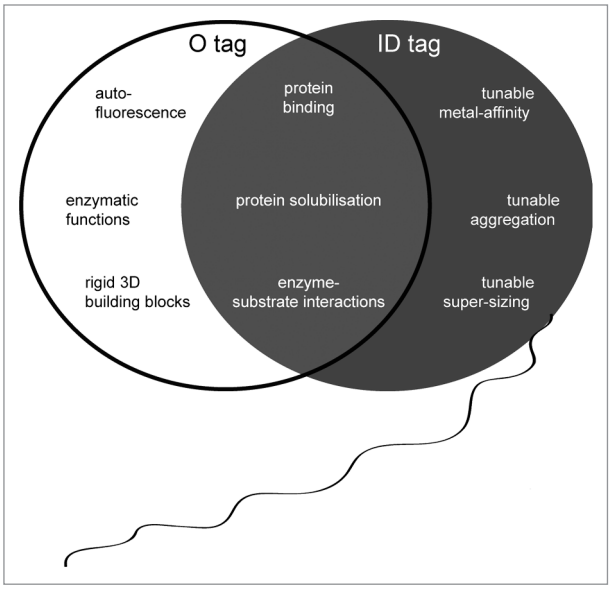

Figure 2. ID tags complement and expand the use of folded tags. Several functions are unique to folded tags or unfolded ID tags. Auto-fluorescence, enzymatic activity and rigidity in 3D are known features of folded proteins. ID tags permit tuning of size enlargement (super-sizing), metal affinity or aggregation-propensity. ID tags and ordered tags (O tags) overlap in many applications including protein purification, protein binding and enzyme-substrate interactions.

Other examples of popular folded tags include enzymes that catalyze the formation of specific covalent bonds. Inteins, for instance, catalyze bond-formation and self-excision.[34] Two halves of self-excising split-intein tags can be appended to both ends of a target peptide to cyclize peptides in vivo for rapid generation of stabilized libraries of small peptides.[35-37] Two recently established enzyme tags called SnapTag and HaloTag facilitate rapid fluorescent labeling.[38] Fluorescent tracing of low-abundant receptors using a rapid but non-denaturing SnapTag coupling protocol was recently demonstrated.[39,40] HaloTag coupling was used to site-specifically couple single protein molecules for optical tweezers studies of protein folding.[41,42]

Small ID Tags Are Popular Multi-Talents

Many of the most commonly used tags are small and predicted to lack structure (**Table 1**). Despite being around 20 residues at maximum, they mediate highly specific and diverse functions, from tunable affinity probes to covalent coupling points.

Scientists can now choose from a large range of small tags, which ensure that the protein fused to it can be highly selectively detected and enriched by specific binders. Frequently used examples for applications in pull-downs and chromatographic purifications include Hexahistidine (short: His6) tags and the 8 amino acids short Strep-tag II.[20,21,43,44] His_6 tags mediate reversible and tunable affinity under a wide range of conditions from near physiological to protein denaturing conditions, have been used successfully in thousands of protein purifications, and are an integral part of most structural biology pipelines. Many small His_6 tagged proteins are sufficiently overproduced in *E. coli* to achieve greater than 95% purity in one immobilised metal affinity chromatography (IMAC) purification step. In purifications from mammalian hosts, histidine-rich metal-binding intracellular proteins might interfere with IMAC. The more selective Strep-tag II can be a valuable alternative or additional purification step along with His_6 tags. Its specificity was further increased by repeats allowing immobilisation of single molecules under applied force (up to 65 pN) for hours and purification of low abundant mammalian proteins.[45-47]

Popular epitope tags for immunoblots, pulldowns, or immunoprecipitations include FLAG tag, Myc tag, V5 tag, and HA tag.[18,48] Since their sequence composition and hydrophobicity varies significantly, screening several variants in parallel helps optimising the results. Repeats often help to increase the effective binding strength or apparent specificity of binding of the epitope tag to its cognate antibody.[49,50] The recently described NSS tag was suggested to be comparable to multi-Myc tags and might be an interesting alternative for future co-immunoprecipitation and protein interaction studies.[51]

The AviTag, or biotin acceptor peptide (BAP), allows for in vivo biotinylation.[38] Biotin has an ultra-high affinity for Streptavidin (K_D of ~10^{-14} M) and forms an unusually stable complex that is still 50% folded at 112°C.[52,53] Many sensitive detection assays, immobilization strategies and nano-patterning strategies exploit the strength of this interaction.[54-56] Recently, "Traptavidin," an engineered variant of streptavidin, further improved the thermostability of the complex and decreased the off-rate by an order of magnitude.[57]

Several small tags have been developed to allow site-specific covalent coupling, either spontaneously or upon addition of a highly specific enzyme. Sortase-mediated protein ligation (SMPL), for instance, is a method to couple polypeptide chains to a wide variety of (bio)molecules including nucleotides, sugars, lipids, organic and inorganic particles, or to circularise peptides efficiently.[58,59] Polypeptides with a C-terminal Sortag (LPXTG motif) can be specifically recognized by Sortase A and transferred to either the natural nucleophile (GGGGG) or triple-Glycine-modified target molecules. Most applications use Staphylococcus aureus Sortase A.[60,61] Complementary semi-orthogonal calcium-independent Sortases allow incorporation of a second coupling point using the LPXTA motif in the same polypeptide chain or enable transpeptidations in vivo in specific cellular compartments.[62-64] The small size of the Sortag, combined with good expression of the Sortase A enzyme as well as the flexibility of Sortase A as "molecular stapler" for a large range of bio-conjugation challenges made this system popular.[65,66]

A major limitation of the Sortase A coupling chemistry is the low efficiency of the reaction at nanomolar or lower concentrations of reactants.[60] Directed evolution of rate-enhanced variants of Sortase A and immobilization of the Sortase A to a flexible solid helped to increase yields and speed of reaction.[67-69] The 6 residues longer ybbR tag can be specifically recognized by the Sfp enzyme.[70] This tag was successfully applied to attach single stalled ribosomes to beads to assess ribosome-associated folding

Table 1. Overview of protein fusion tags and their PONDR-FIT predicted ID propensity

ID [%]	Name	Application	Obeserved effect	Composition	Residues	References
6	MBP	protein solubility	65% > His_6 tag	X	396	22,26
6	NusA	protein solubility	solubility	X	495	122
6	Halo Tag	coupling	covalent coupling	X	297	123
7	GroEL	solubility	aggregation prevention	X (- W)	548	30
9	DsbA	peptide stabilization	high peptide yield	X	208	124
11	GST	affinity	dimer/affinity	X	200	125
11	I_N	split-intein	induced folding	X(- M)	123	35
15	Fc tag	ProtA/G affinity	high affinity (10 nM)	X	232	126
16	Lipoyl	solubility	solubility	X (- RM)	80	127
18	GFP	fluorescence	tunable fluorescence	X (- CW)	238	128,129
18	SNAP-tag	coupling	in vivo labeling	X	182	39,130
18	GFP11	split-GFP	in vivo complementation	X (- NCQPSWY)	17	129
21	Trx	protein solubility	peptide stabilization	X (- HR)	103	18
21	DnaK	solubility	aggregation prevention	X	638	30
22	CBD	chitin affinity	high capacity		51	131
24	mtHsp70	solubility	high yields	X(- C)	625	132
40	Sortag	coupling	covalent coupling	LPXTG	5	58,59,62
41	I_C	split-intein	coupled folding/binding	X (- CWMTWY)	34	35
46	ySUMO	protein solubility	peptide stabilization	X (- CW)	100	85,133
47	GB1	solubility	solubility	X(- PHCR)	64	134
50	His_6 tag	affinity	specific binding	H	6	20
100	XTEN	half life	tunable increase	PESTAG	36–1008	25
100	HAP	half life	3-fold increase	SG	100, 200	23
100	PAS	half life	tunable increase	PAS	100–600	24
100	EB60A	solubility	75% > His_6 tag	PESQ	60	22
100	EB60B	solubility	75% > His_6 tag	PEGQ	60	22
100	EB144	solubility	95% > His_6 tag	PESDGQ	144	22
100	EB250	solubility	100% > His_6 tag	PESDGQMVILF	250	22
100	Antigen 13 rep.	half-life	2-fold increase	PESAK	280	135
100	SAPA repeats	half life	5-fold increase	PASTDH	156	135
100	HRM	half life	4.5–6-fold increase	PASTD	105	136
100	poly(E)	drug solubility	tumors reduced	E	277	137,138
100	ELP	drug efficacy	enhanced efficacy	PVGX	160- 750	139

Table 1. Overview of protein fusion tags and their PONDR-FIT predicted ID propensity (**continued**)						
ID [%]	**Name**	**Application**	**Obeserved effect**	**Composition**	**Residues**	References
100	ELP	half life	8.7h half-life	PVGX	450–600	140
100	Random coil	length	tunable gel	PEGQAN	100–800	141
100	NNT/NNS	drug properties	tuned glycosylation	NTS	60–750	142
100	ELP	purification	phase cycling possible	PVGX	50–900	114
100	ELP	solubility of Ig	improved solubility	PESTAGX(- HY)	22–61	143
100	ELP	cost reduction	tailored transition	PVGX	450	144
100	ELP	simple purification	effective as His tag	PVGX	?	145
100	His_3	immunoaffinity	co-crystallization	H	3	74
100	c-MYC	Detection	highly specific	ELDQIKS	10	18
100	NSS	Detection	highly specific	FINQHMKT	9	51
100	Strep-tag II	streptactin affinity	highly specific	WSHPQFEK	8	43,44
100	V5 tag	detection	highly specific	PSTKLINDG	14	146
100	HA tag	detection	highly specific	YPDAV	9	147
100	FLAG tag	detection	highly specific	DKY	8	18
100	Avi/BAP tag	biotin binding	biotinylation by BirA	EIANDQGHLKF	15	38
100	Ybbr tag	coupling	covalent coupling	ALSDEIKF	11	70
100	SpyTag	coupling	covalent coupling	KAVDIHPTYM	13	72,73
100	Zn hook	dimerization	fM dimer	CGKRADELPTV	14	148
100	Leucine Zipper	dimerization	nM hetero-dimer	KELQANW	30+30	149,150
100	Arg tag	purification	surface adhesion	R	5	75,151
100	tetracystein tag	FlAsH binding	FlAsH dequench	CPG	6	152
100	CBP	calmodulin affinity	high affinity	KASRNIFGLWV	26	83,153
n.a.	Cysteine	specific reactivity	chemical coupling	C	1	154

Composition is indicated by single-letter representation of amino acids with X denoting any of the 20 amino acids; amino acids in brackets are not occurring in the respective tag. For calculation of the indicated ID fractions in percent of the total number of residues, we counted residues with a PONDR-FIT predicted disorder propensity > 0.5 as disordered and residues with lower predicted disorder propensity as ordered.

in optical tweezer studies.[71] The recently engineered Isopeptag and Spytag permit enzyme-free, efficient site-specific covalent coupling to the spontaneously bond-forming proteins Spy0128 and SpyCatcher and might become a valuable new addition to the protein engineering toolbox.[72,73]

The Better Tag Remains Unseen

Folded tags typically need to be removed after protein purification to prevent structural or functional interference during biophysical or structural characterization of the target protein. Crystallographic structure determination usually even requires removal of short disordered tags as their flexibility may hamper crystallization, although a few cases have been reported in which the presence of the Arg-6, His-3, His-6, FLAG, or c-Myc tag did not preclude structural studies.[74-77] Several highly site-specific enzymes cleave at short but rather rare sequence motifs, which lack 3D structure propensity, and are inserted between the tag and protein of interest: ENLYFQ*G/S for TEV protease, DDDK* for Enterokinase, IE/DGR* for Factor Xa, LEVLFQ* GP for HRV 3C, LPXT*G for Sortase A, LVPR* GS for Thrombin (cleavage occurs at *).[78-81] Sortase A uses the same motif for both specific cleavage and subsequent protein ligation. Thus, tag removal can be simultaneously combined with

attachment of other tags using Sortase A or inteins, for instance to include NMR-invisible solubility tags.[82,83]

Although the above-mentioned cleavage motifs are relatively rare in proteomes, it is not excluded that they are present in some natural or engineered target proteins. Hence selecting a unique cleavage site to specifically remove a fusion protein from the protein of interest requires caution. Enhanced cleavage specificity can be attained if cleavage depends on recognition of a unique 3D structure as in the case of Ubiquitin, SUMO, or Intein proteins.[84-87] Briefly, "visible," folded tags can be efficiently removed using unstructured motifs and "invisible," either unlabelled or unstructured tags are often compatible with or even required for downstream applications.

ID Tags Are Not Limited by the Need to Fold During Purification

Structured tags are often structurally well-characterized by high-resolution methods like X-ray crystallography and NMR spectroscopy (**Fig. 1**). Because of their defined, compact structure, they are less likely to be degraded by proteases, and amenable to rational design and engineering of their thermal stability.[88-91] However, if the conditions are not ideal for folding, for instance upon exposure to physicochemical stresses or when folding intermediates of the target protein engage in non-native interactions, irreversible aggregation and loss or unpredictable alteration of tag function could occur.[92] Even though folded tags are well-established, their relatively large size and their need to fold can thus be limiting for many applications.

IDP tags and short ID tags are structurally less characterized, newer, and therefore less established for some applications. Their larger solvent-exposed surface and lack of rigid structure might make them more susceptible to proteolysis. Several variants of large ID tags have, however, been described, which can be expressed at high yields for biophysical characterization.[24,93] Up to 4000 mg/L of large ID tag fusions have been obtained in a recent example.[94] ID tags do not need to fold, therefore the stability of ID tags in denaturing conditions is frequently exploited in denaturing affinity purifications.[20]

ID and Its Flavors Can be Predicted With High Accuracy

Propensity for ID is encoded in the sequence of amino acids and can be efficiently predicted for all sequenced proteomes.[1] Most prediction algorithms combine physico-chemical properties of amino acids and empirical knowledge of structure propensities as derived from large experimental data sets.[95] One of the most widely used algorithm is PONDR. Its accuracy is, like similar methods, around 80% for water-soluble proteins; these estimates might be biased by the limited availability of data for large disordered regions above 40 residues, which rarely crystallize.[4,96-98] Thousands of sequenced genomes have been analyzed with this algorithm. These PONDR predictions indicate that some 40% of all eukaryotic protein residues are located in ID regions.[1] More recent methods of ID prediction combine several predictors into 1 meta-prediction, which significantly improved the accuracy of predictions in several cases.[99] Disorder predictions are increasingly accurate and fast.[98] Future progress in algorithm development will benefit from novel large-scale experimental data sets on ID propensity in larger fractions of complete proteomes including yet understudied membrane proteins.[14,90,100,101]

A good experimental proxy for the average apparent size and extension of an IDP is the hydrodynamic radius.[102-104] Both hydrodynamic experiments (such as SEC or more recently SEC-MALLS) as well as scattering experiments (DLS, SAXS, SANS) can give insights into hydration radii.[103,104] Drawbacks of these experimental approaches are that they require either substantial amounts of highly purified protein or access to expensive equipment. Computational approaches to model global features of IDPs are not limited by these restraints. An empirically optimised formula to predict hydration radii has been proposed based on a large set of experimentally solved radii and statistical correlation with sequence features of these proteins.[105] Mainly based on proline content and net charges, this predictor achieves high accuracy for many small disordered proteins suggesting broad applicability.[105] An alternative approach approximates IDPs as polyampholytes because of their enrichment in both positive and negative charges. While weak polyampholytes collapse to globule-like shapes, stronger polyampholytes are more extended and their shape depends on local distribution of positive and negative charges within the primary sequence. Their predicted ensemble properties can be efficiently modeled computationally using implicit salvation in the Absinth force-field.[106] Interactions of small ID regions with folded proteins often involve coupled folding and binding.[107-109] Combining initial rigid-body docking and subsequent flexible adjustment appears to be a computationally efficient solution to this problem.[110] Several novel computational approaches expand the scope of prediction of ensemble properties of IDPs.

Can we also predict aggregation propensity of ID regions from their amino acid sequence? Irreversible aggregation is a well-known disease-associated feature of multiple naturally occurring IDPs.[111] Several ID regions are, however, highly soluble or aggregate reversibly. Sequence composition, particularly the relative fractions of proline and glycine, strongly affect amyloidogenicity of unfolded polypeptides.[112] Also a large percentage of charges, especially glutamate, can help preventing aggregation in a broad range of conditions. Several net negatively charged proteins like tau or the engineered PESTAG sequences remain soluble after heating to 95°C.[93,113] Precise control of reversible thermal aggregation has been demonstrated using engineered elastin-like polymers (ELPs).[114] ELPs that contain large fractions of proline, glycine, valine enable cycles of thermally controlled reversible aggregation and solubilisation around an engineered transition temperature between 40°C and 80°C.[114] Thus, multiple strategies help predicting and controlling the aggregation-propensity of IDPs.

Designed, Large ID Tags: Applications from Biotechnology to Next Generation Protein Pharmaceuticals

Increased understanding of naturally occurring IDPs and improved predictability of their features enabled the design of several ID tags for overcoming protein aggregation, facilitating non-chromatographic purification and tuning pharmaco-kinetic profiles of protein drugs.

Inspired by the amino acid composition of plant dehydrins, which efficiently cope with extremes of dessication stress and prevent aggregation upon rehydration,[115] and aided by predictions of ID and solubility, the Keith Dunker lab recently developed several large, negatively charged, fully disordered "entropic bristle" (EB) tags to solubilize difficult-to-produce proteins. EB fusions successfully solubilized target proteins from several protein classes including kinases, transcription factors, proteases, and neurodegeneration-related proteins.[22] The EB tags did not interfere with GST enzyme function and structural stability and therefore did not have to be removed after production in this particular case.[22] If tag removal is, however, required and the target protein is stably folded and proteolytically resistant, rapid thermally accelerated proteolysis of the unstructured tags could be performed using the inexpensive, commercially available enzyme thermolysin.[90]

Many proteins have been purified using short affinity tags and chromatography.[20] Scaling up chromatography-based purifications can be very costly as most chromatography materials are costly. Chromatography-free purification methods that yield highly purified proteins are therefore desirable for many applications. One possible solution to this problem exploits repeated cycles of reversible aggregation and solubilisation of designed elastin-like sequences, which have a defined transition temperature point.[116] A recent extension of this concept by simultaneous genetic fusion of both an elastin-like peptide (ELP) sequence and a Sortase A fusion tag achieved one-step purification of large amounts of soluble target proteins without the need of chromatography for several tested proteins.[117] This method yielded, for instance, 28 mg/L pure, sortagged GFP and 35 mg/L pure, sortagged Trx without chromatography.[117]

Protein and peptide pharmaceuticals have long been suggested as possible complementation for traditional small-molecule drugs and inspire hopes for reduced side-effects and more targeted therapies. Current limitations for their broader application include large production costs, less convenient administration by injection and a short half-life. The plasma half-lives of peptide pharmaceuticals can be as short as 10 min.[93] Increasing the half-life could help overcoming both cost problems and simplify drug dosage and application. Several groups have recently described complementary, effective strategies to use designed ID tags for extending the half-life of protein pharmaceuticals. The Pim Stemmer lab developed super-sized, unfolded PESTAG sequences, which increased the half-time of peptides and hormones up to 125-fold by mimicking attachments of globular molecules in the Megadalton range.[118] Curiously, this supercharged fusion also completely prevented the thermal aggregation of human growth hormone and might be beneficial in transport and storage of protein drugs. The Arne Skerra lab more recently presented the PASylation technology,[24] which uses an uncharged but stably unfolded and conformationally extended polymer consisting of proline, alanine, and serine. This hydrophilic polypeptide shows remarkably similar biophysical behavior to polyethylene glycol (PEG) while offering biodegradability. PASylation mediates tunable large increase in plasma half-life of protein pharmaceuticals and its fusions can be functionally expressed at very high levels (4000 mg/L *E.coli* fermentation culture volume).[94] Thus, these novel, large ID fusion tags might become invaluable alternatives for organic polymer extensions using polyethylene glycol that turned out to be more immunogenic than previously anticipated.[119]

Conclusions

Folded and ID tags can have overlapping as well as different features, resulting in complementarity and synergisms of their applications (**Fig. 2**):

Folded tags excel in the following features: (1) Rigidity in 3D (2) Specific enzymatic functions (3) Auto-fluorescence

ID tags, on the other hand, permit tuning: (1) enlargement ("super-sizing") compared with compact globular proteins (2) metal affinity from rapidly reversible binding (His_6 tag) up to femtomolar affinity (Zn hook) (3) thermal aggregation behavior

ID tags and folded/ordered tags overlap in: (1) Protein solubilization (2) Protein binding (reversible to covalent) (3) Enzyme-substrate interactions

In a nutshell, ID tags are invaluable additions to the toolbox of protein sciences. Future applications of ID tags will be catalyzed by the rapid growth of IDP research.[16,102,103] In nature, ID tails have evolutionary diverged into multifarious roles.[120,121] We can learn from nature's great example and create novel ID tails to harness their unique features for new purposes.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We are grateful to Keith Dunker, PhD, Fatemeh Moayed, Nuria Taberner, Dr. Stefan Rüdiger and Md. Arif Kamal for critical reading and fruitful discussions on earlier versions of the manuscript. We thank Professor Dr. Arne Skerra, Technische Universität München, for a picture of his modelled PASylated Fab fragment which illustrates the emerging concept of large intrinsically disordered fusion tags to improve the pharmacological properties of therapeutic proteins. This graphics is used as cover art of Intrinsically Disordered Proteins in 2014.

References

1. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 2012; 30:137-49; PMID:22702725; http://dx.doi.org/10.1080/07391102.2012.675145
2. Petsalaki E, Russell RB. Peptide-mediated interactions in biological systems: new discoveries and applications. Curr Opin Biotechnol 2008; 19:344-50; PMID:18602004; http://dx.doi.org/10.1016/j.copbio.2008.06.004
3. Uversky VN. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. FEBS Lett 2013; 587:1891-901; PMID:23665034; http://dx.doi.org/10.1016/j.febslet.2013.04.042
4. Xing Y, Takemaru K, Liu J, Berndt JD, Zheng JJ, Moon RT, Xu W. Crystal structure of a full-length beta-catenin. Structure 2008; 16:478-87; PMID:18334222; http://dx.doi.org/10.1016/j.str.2007.12.021
5. Wayne N, Bolon DN. Charge-rich regions modulate the anti-aggregation activity of Hsp90. J Mol Biol 2010; 401:931-9; PMID:20615417; http://dx.doi.org/10.1016/j.jmb.2010.06.066
6. Minde DP, Anvarian Z, Rüdiger SG, Maurice MM. Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer? Mol Cancer 2011; 10:101; PMID:21859464; http://dx.doi.org/10.1186/1476-4598-10-101
7. Noutsou M, Duarte AMS, Anvarian Z, Didenko T, Minde DP, Kuper I, de Ridder I, Oikonomou C, Friedler A, Boelens R, et al. Critical scaffolding regions of the tumor suppressor Axin1 are natively unfolded. J Mol Biol 2011; 405:773-86; PMID:21087614; http://dx.doi.org/10.1016/j.jmb.2010.11.013
8. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, et al. DisProt: the Database of Disordered Proteins. Nucleic Acids Res 2007; 35:D786-93; PMID:17145717; http://dx.doi.org/10.1093/nar/gkl893 .
9. Quint S, Widmaier S, Minde D, Hornburg D, Langosch D, Scharnagl C. Residue-specific side-chain packing determines the backbone dynamics of transmembrane model helices. Biophys J 2010; 99:2541-9; PMID:20959095; http://dx.doi.org/10.1016/j.bpj.2010.08.031
10. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004; 32:1037-49; PMID:14960716; http://dx.doi.org/10.1093/nar/gkh253 .
11. Minde DP, Radli M, Forneris F, Maurice MM, Rüdiger SGD. Large extent of disorder in Adenomatous Polyposis Coli offers a strategy to guard Wnt signalling against point mutations. PLoS ONE 2013; 8:e77257; PMID:24130866; http://dx.doi.org/10.1371/journal.pone.0077257 .
12. Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 2011; 39:D253-60; PMID:21081558; http://dx.doi.org/10.1093/nar/gkq1159
13. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. Nature 2013; 498:390-4; PMID:23783631; http://dx.doi.org/10.1038/nature12294
14. Tsvetkov P, Asher G, Paz A, Reuven N, Sussman JL, Silman I, Shaul Y. Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. Proteins 2008; 70:1357-66; PMID:17879262; http://dx.doi.org/10.1002/prot.21614
15. Elbaum-Garfinkle S, Ramlall T, Rhoades E. The role of the lipid bilayer in tau aggregation. Biophys J 2010; 98:2722-30; PMID:20513417; http://dx.doi.org/10.1016/j.bpj.2010.03.013
16. Uversky VN. Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol 2011; 43:1090-103; PMID:21501695; http://dx.doi.org/10.1016/j.biocel.2011.04.001
17. Ford CF, Suominen I, Glatz CE. Fusion tails for the recovery and purification of recombinant proteins. Protein Expr Purif 1991; 2:95-107; PMID:1821793; http://dx.doi.org/10.1016/1046-5928(91)90057-P
18. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2003; 60:523-33; PMID:12536251; http://dx.doi.org/10.1007/s00253-002-1158-6
19. Kohl T, Schmidt C, Wiemann S, Poustka A, Korf U. Automated production of recombinant human proteins as resource for proteome research. Proteome Sci 2008; 6:4; PMID:18226205; http://dx.doi.org/10.1186/1477-5956-6-4
20. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, et al.; Structural Genomics Consortium; China Structural Genomics Consortium; Northeast Structural Genomics Consortium. Protein production and purification. Nat Methods 2008; 5:135-46; PMID:18235434; http://dx.doi.org/10.1038/nmeth.f.202
21. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, et al. Large-scale identification of protein-protein interaction of Escherichia coli K-12. Genome Res 2006; 16:686-91; PMID:16606699; http://dx.doi.org/10.1101/gr.4527806
22. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van YY, Dunker AK. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry 2012; 51:7250-62; PMID:22924672; http://dx.doi.org/10.1021/bi300653m
23. Schlapschy M, Theobald I, Mack H, Schottelius M, Wester H-J, Skerra A. Fusion of a recombinant antibody fragment with a homo-amino-acid polymer: effects on biophysical properties and prolonged plasma half-life. Protein Eng Des Sel 2007; 20:273-84; PMID:17595342; http://dx.doi.org/10.1093/protein/gzm020
24. Schlapschy M, Binder U, Börger C, Theobald I, Wachinger K, Kisling S, Haller D, Skerra A. PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. Protein Eng Des Sel 2013; 26:489-501; PMID:23754528; http://dx.doi.org/10.1093/protein/gzt023
25. Geething NC, To W, Spink BJ, Scholle MD, Wang CW, Yin Y, Yao Y, Schellenberger V, Cleland JL, Stemmer WPC, et al. Gcg-XTEN: an improved glucagon capable of preventing hypoglycemia without increasing baseline blood glucose. PLoS One 2010; 5:e10175; PMID:20418955; http://dx.doi.org/10.1371/journal.pone.0010175
26. Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999; 8:1668-74; PMID:10452611; http://dx.doi.org/10.1110/ps.8.8.1668
27. Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 1999; 65:382-8; PMID:10506413; http://dx.doi.org/10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
28. Brandt T, Kaar JL, Fersht AR, Veprintsev DB. Stability of p53 homologs. PLoS One 2012; 7:e47889; PMID:23112865; http://dx.doi.org/10.1371/journal.pone.0047889
29. Japrung D, Dogan J, Freedman KJ, Nadzeyka A, Bauerdick S, Albrecht T, Kim MJ, Jemth P, Edel JB. Single-molecule studies of intrinsically disordered proteins using solid-state nanopores. Anal Chem 2013; 85:2449-56; PMID:23327569; http://dx.doi.org/10.1021/ac3035025.
30. Kyratsous CA, Silverstein SJ, DeLong CR, Panagiotidis CA. Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in Escherichia coli. Gene 2009; 440:9-15; PMID:19328840; http://dx.doi.org/10.1016/j.gene.2009.03.011
31. Tsien RY. The 2009 Lindau Nobel Laureate Meeting: Roger Y. Tsien, Chemistry 2008. J Vis Exp 2010; 29:2396; PMID:20072108.
32. Van der Zand A, Gent J, Braakman I, Tabak HF. Biochemically distinct vesicles from the endoplasmic reticulum fuse to form peroxisomes. Cell 2012; 149:397–409.
33. van der Zand A, Tabak HF. Peroxisomes: offshoots of the ER. Curr Opin Cell Biol 2013; 25:449-54; PMID:23773570; http://dx.doi.org/10.1016/j.ceb.2013.05.004
34. Paulus H. Inteins as enzymes. Bioorg Chem 2001; 29:119-29; PMID:11437387; http://dx.doi.org/10.1006/bioo.2001.1203 .
35. Zheng Y, Wu Q, Wang C, Xu M-Q, Liu Y. Mutual synergistic protein folding in split intein. Biosci Rep 2012; 32:433-42; PMID:22681309; http://dx.doi.org/10.1042/BSR20120049
36. Ramirez M, Valdes N, Guan D, Chen Z. Engineering split intein DnaE from Nostoc punctiforme for rapid protein purification. Protein Eng Des Sel 2013; 26:215-23; ; PMID:23223807; http://dx.doi.org/10.1093/protein/gzs097
37. Tavassoli A, Benkovic SJ. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. Nat Protoc 2007; 2:1126-33; PMID:17546003; http://dx.doi.org/10.1038/nprot.2007.152
38. Tirat A, Freuler F, Stettler T, Mayr LM, Leder L. Evaluation of two novel tag-based labelling technologies for site-specific modification of proteins. Int J Biol Macromol 2006; 39:66-76; PMID:16503347; http://dx.doi.org/10.1016/j.ijbiomac.2006.01.012
39. Koo B-K, Spit M, Jordens I, Low TY, Stange DE, van de Wetering M, van Es JH, Mohammed S, Heck AJR, Maurice MM, et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. Nature 2012; 488:665-9; PMID:22895187; http://dx.doi.org/10.1038/nature11308
40. Galbraith MD, Allen MA, Bensard CL, Wang X, Schwinn MK, Qin B, Long HW, Daniels DL, Hahn WC, Dowell RD, et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell 2013; 153:1327-39; PMID:23746844; http://dx.doi.org/10.1016/j.cell.2013.04.048
41. Fernández-Suárez M, Ting AY. Fluorescent probes for super-resolution imaging in living cells. Nat Rev Mol Cell Biol 2008; 9:929-43; PMID:19002208; http://dx.doi.org/10.1038/nrm2531
42. Aubin-Tam M-E, Olivares AO, Sauer RT, Baker TA, Lang MJ. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. Cell 2011; 145:257-67; PMID:21496645; http://dx.doi.org/10.1016/j.cell.2011.03.036
43. Schmidt TG, Koepke J, Frank R, Skerra A. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. J Mol Biol 1996; 255:753-66 ; http://dx.doi.org/10.1006/jmbi.1996.0061

44. Voss S, Skerra A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein Eng 1997; 10:975-82; PMID:9415448; http://dx.doi.org/10.1093/protein/10.8.975
45. Busby M, Stadler LKJ, Ko Ferrigno P, Davis JJ. Optimisation of a multivalent Strep tag for protein detection. Biophys Chem 2010; 152:170-7; PMID:20970240; http://dx.doi.org/10.1016/j.bpc.2010.09.005
46. Moayed F, Mashaghi A, Tans SJ. A polypeptide-DNA hybrid with selective linking capability applied to single molecule nano-mechanical measurements using optical tweezers. PLoS One 2013; 8:e54440; PMID:23336001; http://dx.doi.org/10.1371/journal.pone.0054440
47. Halff EF, Diebolder CA, Versteeg M, Schouten A, Brondijk THC, Huizinga EG. Formation and structure of a NAIP5-NLRC4 inflammasome induced by direct interactions with conserved N- and C-terminal regions of flagellin. J Biol Chem 2012; 287:38460-72; PMID:23012363; http://dx.doi.org/10.1074/jbc.M112.393512
48. Fritze CE, Anderson TR. Epitope tagging: general method for tracking recombinant proteins. Methods Enzymol 2000; 327:3-16; PMID:11044970; http://dx.doi.org/10.1016/S0076-6879(00)27263-7
49. Jakobi AJ, Mashaghi A, Tans SJ, Huizinga EG. Calcium modulates force sensing by the von Willebrand factor A2 domain. Nat Commun 2011; 2:385; PMID:21750539; http://dx.doi.org/10.1038/ncomms1385
50. Mashaghi A, Kramer G, Bechtluft P, Zachmann-Brand B, Driessen AJM, Bukau B, Tans SJ. Reshaping of the conformational search of a protein by the chaperone trigger factor. Nature 2013; 500:98-101; PMID:23831649; http://dx.doi.org/10.1038/nature12293
51. Cheng HW, Chen KC, Raja JAJ, Li JX, Yeh SD. An efficient tag derived from the common epitope of tospoviral NSs proteins for monitoring recombinant proteins expressed in both bacterial and plant systems. J Biotechnol 2013; 164:510-9; PMID:23403362; http://dx.doi.org/10.1016/j.jbiotec.2013.02.002
52. Weber PC, Ohlendorf DH, Wendoloski JJ, Salemme FR. Structural origins of high-affinity biotin binding to streptavidin. Science 1989; 243:85-8; PMID:2911722; http://dx.doi.org/10.1126/science.2911722
53. González M, Argaraña CE, Fidelio GD. Extremely high thermal stability of streptavidin and avidin upon biotin binding. Biomol Eng 1999; 16:67-72; PMID:10796986; http://dx.doi.org/10.1016/S1050-3862(99)00041-8
54. Chu YW, Wang BY, Lin H-S, Lin T-Y, Hung Y-J, Engebretson DA, Lee W, Carey JR. Layer by layer assembly of biotinylated protein networks for signal amplification. Chem Commun (Camb) 2013; 49:2397-9; PMID:23329177; http://dx.doi.org/10.1039/c2cc38233d
55. Heller I, Sitters G, Broekmans OD, Farge G, Menges C, Wende W, Hell SW, Peterman EJG, Wuite GJL. STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. Nat Methods 2013; 10:910-6; PMID:23934077; http://dx.doi.org/10.1038/nmeth.2599
56. Kay BK, Thai S, Volgina VV. High-throughput biotinylation of proteins. In: Doyle SA, editor. High Throughput Protein Expression and Purification. Totowa, NJ: Springer; 2009; 185–96.
57. Chivers CE, Crozat E, Chu C, Moy VT, Sherratt DJ, Howarth M. A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat Methods 2010; 7:391-3; PMID:20383133; http://dx.doi.org/10.1038/nmeth.1450
58. Mao H, Hart SA, Schink A, Pollok BA. Sortase-mediated protein ligation: a new method for protein engineering. J Am Chem Soc 2004; 126:2670-1; PMID:14995162; http://dx.doi.org/10.1021/ja039915e
59. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL. Sortagging: a versatile method for protein labeling. Nat Chem Biol 2007; 3:707-8; PMID:17891153; http://dx.doi.org/10.1038/nchembio.2007.31
60. Levary DA, Parthasarathy R, Boder ET, Ackerman ME. Protein-protein fusion catalyzed by sortase A. PLoS One 2011; 6:e18342; PMID:21494692; http://dx.doi.org/10.1371/journal.pone.0018342
61. Parthasarathy R, Subramanian S, Boder ET. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. Bioconjug Chem 2007; 18:469-76; PMID:17302384; http://dx.doi.org/10.1021/bc060339w
62. Theile CS, Witte MD, Blom AEM, Kundrat L, Ploegh HL, Guimaraes CP. Site-specific N-terminal labeling of proteins using sortase-mediated reactions. Nat Protoc 2013; 8:1800-7; PMID:23989674; http://dx.doi.org/10.1038/nprot.2013.102
63. Witte MD, Cragnolini JJ, Dougan SK, Yoder NC, Popp MW, Ploegh HL. Preparation of unnatural N-to-N and C-to-C protein fusions. Proc Natl Acad Sci U S A 2012; 109:11993-8; PMID:22778432; http://dx.doi.org/10.1073/pnas.1205427109
64. Strijbis K, Spooner E, Ploegh HL. Protein ligation in living cells using sortase. Traffic 2012; 13:780-9; PMID:22348280; http://dx.doi.org/10.1111/j.1600-0854.2012.01345.x
65. Clow F, Fraser JD, Proft T. Immobilization of proteins to biacore sensor chips using Staphylococcus aureus sortase A. Biotechnol Lett 2008; 30:1603-7; PMID:18414796; http://dx.doi.org/10.1007/s10529-008-9718-1
66. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. Proc Natl Acad Sci U S A 1999; 96:12424-9; PMID:10535938; http://dx.doi.org/10.1073/pnas.96.22.12424
67. Chan L, Cross HF, She JK, Cavalli G, Martins HFP, Neylon C. Covalent attachment of proteins to solid supports and surfaces via Sortase-mediated ligation. PLoS One 2007; 2:e1164; PMID:18000537; http://dx.doi.org/10.1371/journal.pone.0001164
68. Chen I, Dorr BM, Liu DR. A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A 2011; 108:11399-404; PMID:21697512; http://dx.doi.org/10.1073/pnas.1101046108
69. Steinhagen M, Zunker K, Nordsieck K, Beck-Sickinger AG. Large scale modification of biomolecules using immobilized sortase A from Staphylococcus aureus. Bioorg Med Chem 2013; 21:3504-10; PMID:23598248; http://dx.doi.org/10.1016/j.bmc.2013.03.039
70. Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc Natl Acad Sci U S A 2005; 102:15815-20; PMID:16236721; http://dx.doi.org/10.1073/pnas.0507705102
71. Kaiser CM, Goldman DH, Chodera JD, Tinoco I Jr., Bustamante C. The ribosome modulates nascent protein folding. Science 2011; 334:1723-7; PMID:22194581; http://dx.doi.org/10.1126/science.1209740
72. Abe H, Wakabayashi R, Yonemura H, Yamada S, Goto M, Kamiya N. Split Spy0128 as a potent scaffold for protein cross-linking and immobilization. Bioconjug Chem 2013; 24:242-50; PMID:23350748; http://dx.doi.org/10.1021/bc300606b
73. Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci USA 2012; 109:E690-7; PMID:22366317; http://dx.doi.org/10.1073/pnas.1115485109
74. Kaufmann M, Lindner P, Honegger A, Blank K, Tschopp M, Capitani G, Plückthun A, Grütter MG. Crystal structure of the anti-His tag antibody 3D5 single-chain fragment complexed to its antigen. J Mol Biol 2002; 318:135-47; PMID:12054774; http://dx.doi.org/10.1016/S0022-2836(02)00038-4
75. Bucher MH, Evdokimov AG, Waugh DS. Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. Acta Crystallogr D Biol Crystallogr 2002; 58:392-7; PMID:11856823; http://dx.doi.org/10.1107/S0907444901021187
76. Håkansson K, Broder D, Wang AH-J, Miller CG. Crystallization of peptidase T from Salmonella typhimurium. Acta Crystallogr D Biol Crystallogr 2000; 56:924-6; PMID:10930847; http://dx.doi.org/10.1107/S0907444900006375
77. Garrett TPJ, McKern NM, Lou M, Frenkel MJ, Bentley JD, Lovrecz GO, Elleman TC, Cosgrove LJ, Ward CW. Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. Nature 1998; 394:395-9; PMID:9690478; http://dx.doi.org/10.1038/28668
78. Vergis JM, Wiener MC. The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. Protein Expr Purif 2011; 78:139-42; PMID:21539919; http://dx.doi.org/10.1016/j.pep.2011.04.011
79. Guan KL, Dixon JE. Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. Anal Biochem 1991; 192:262-7; PMID:1852137; http://dx.doi.org/10.1016/0003-2697(91)90534-Z
80. Sibbald MJJB, Yang X-M, Tsompanidou E, Qu D, Hecker M, Becher D, Buist G, van Dijl JM. Partially overlapping substrate specificities of staphylococcal group A sortases. Proteomics 2012; 12:3049-62; PMID:22930668; http://dx.doi.org/10.1002/pmic.201200144
81. Kapust RB, Tözsér J, Copeland TD, Waugh DS. The P1¢ specificity of tobacco etch virus protease. Biochem Biophys Res Commun 2002; 294:949-55; PMID:12074568; http://dx.doi.org/10.1016/S0006-291X(02)00574-0
82. Kobashigawa Y, Kumeta H, Ogura K, Inagaki F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. J Biomol NMR 2009; 43:145-50; PMID:19140010; http://dx.doi.org/10.1007/s10858-008-9296-5
83. Durst FG, Ou HD, Löhr F, Dötsch V, Straub WE. The better tag remains unseen. J Am Chem Soc 2008; 130:14932-3; PMID:18937478; http://dx.doi.org/10.1021/ja806212j
84. Lee EK, Hwang JH, Shin DY, Kim DI, Yoo YJ. Production of recombinant amyloid-beta peptide 42 as an ubiquitin extension. Protein Expr Purif 2005; 40:183-9; PMID:15721787; http://dx.doi.org/10.1016/j.pep.2004.12.014
85. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics 2004; 5:75-86; PMID:15263846; http://dx.doi.org/10.1023/B:JSFG.0000029237.70316.52
86. Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 2001; 29:3757-74; PMID:11557808; http://dx.doi.org/10.1093/nar/29.18.3757

87. Karagöz GE, Sinnige T, Hsieh O, Rüdiger SGD. Expressed protein ligation for a large dimeric protein. Protein Eng Des Sel 2011; 24:495-501; PMID:21335434; http://dx.doi.org/10.1093/protein/gzr007
88. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 2006; 24:79-88; PMID:16369541; http://dx.doi.org/10.1038/nbt1172
89. Cabantous S, Waldo GS. In vivo and in vitro protein solubility assays using split GFP. Nat Methods 2006; 3:845-54; PMID:16990817; http://dx.doi.org/10.1038/nmeth932
90. Minde DP, Maurice MM, Rüdiger SGD. Determining biophysical protein stability in lysates by a fast proteolysis assay, FASTpp. PLoS One 2012; 7:e46147; PMID:23056252; http://dx.doi.org/10.1371/journal.pone.0046147
91. Nikolova PV, Henckel J, Lane DP, Fersht AR. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proc Natl Acad Sci U S A 1998; 95:14675-80; PMID:9843948; http://dx.doi.org/10.1073/pnas.95.25.14675
92. Evans TC Jr., Xu MQ. Intein-mediated protein ligation: harnessing nature's escape artists. Biopolymers 1999; 51:333-42; PMID:10685044; http://dx.doi.org/10.1002/(SICI)1097-0282(1999)51:5<333::AID-BIP3>3.0.CO;2-#
93. Schellenberger V, Wang C-W, Geething NC, Spink BJ, Campbell A, To W, Scholle MD, Yin Y, Yao Y, Bogin O, et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat Biotechnol 2009; 27:1186-90; PMID:19915550; http://dx.doi.org/10.1038/nbt.1588
94. Di Cesare S, Binder U, Maier T, Skerra A. High-Yield Production of PASylated Human Growth Hormone Using Secretory E. coli Technology. Bioprocess Int 2013; Forthcoming
95. Cozzetto D, Jones DT. The contribution of intrinsic disorder prediction to the elucidation of protein function. Curr Opin Struct Biol 2013; 23:467-72; PMID:23466039; http://dx.doi.org/10.1016/j.sbi.2013.02.001
96. Burnley BT, Afonine PV, Adams PD, Gros P. Modelling dynamics in protein crystal structures by ensemble refinement. Elife 2012; 1:e00311
97. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. Automated identification of functional dynamic contact networks from X-ray crystallography. Nat Methods 2013; 10:896-902; PMID:23913260; http://dx.doi.org/10.1038/nmeth.2592
98. Monastyrskyy B, Fidelis K, Moult J, Tramontano A, Kryshtafovych A. Evaluation of disorder predictions in CASP9. Proteins 2011; 79(Suppl 10):107-18; PMID:21928402; http://dx.doi.org/10.1002/prot.23161
99. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010; 1804:996-1010; PMID:20100603; http://dx.doi.org/10.1016/j.bbapap.2010.01.011
100. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. Nat Biotechnol 2012; 30:1072-80; PMID:23138306; http://dx.doi.org/10.1038/nbt.2419
101. Johnson DE, Xue B, Sickmeier MD, Meng J, Cortese MS, Oldfield CJ, Le Gall T, Dunker AK, Uversky VN. High-throughput characterization of intrinsic disorder in proteins from the Protein Structure Initiative. J Struct Biol 2012; 180:201-15; PMID:22651963; http://dx.doi.org/10.1016/j.jsb.2012.05.013
102. Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 2012; 37:509-16; PMID:22989858; http://dx.doi.org/10.1016/j.tibs.2012.08.004
103. Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 2013; 22:693-724; PMID:23553817; http://dx.doi.org/10.1002/pro.2261
104. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002; 11:739-56; PMID:11910019; http://dx.doi.org/10.1110/ps.4210102
105. Marsh JA, Forman-Kay JD. Sequence determinants of compaction in intrinsically disordered proteins. Biophys J 2010; 98:2383-90; PMID:20483348; http://dx.doi.org/10.1016/j.bpj.2010.02.006
106. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A 2013; 110:13392-7; PMID:23901099; http://dx.doi.org/10.1073/pnas.1304749110
107. Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E, Wang J. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proc Natl Acad Sci U S A 2013; 110:E3743-52; PMID:24043820; http://dx.doi.org/10.1073/pnas.1308381110
108. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 2002; 415:549-53; PMID:11823864; http://dx.doi.org/10.1038/415549a.
109. Lacy ER, Filippov I, Lewis WS, Otieno S, Xiao L, Weiss S, Hengst L, Kriwacki RW. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. Nat Struct Mol Biol 2004; 11:358-64; PMID:15024385; http://dx.doi.org/10.1038/nsmb746
110. Trellet M, Melquiond ASJ, Bonvin AMJJ. A unified conformational selection and induced fit approach to protein-peptide docking. PLoS One 2013; 8:e58769; PMID:23516555; http://dx.doi.org/10.1371/journal.pone.0058769
111. Selkoe DJ. SnapShot: pathobiology of Alzheimer's disease. Cell 2013; 154:468–468.
112. Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. Structure 2006; 14:1667-76; PMID:17098192; http://dx.doi.org/10.1016/j.str.2006.09.008
113. Wang W, Wu X, Xiong E, Tai F. Improving gel-based proteome analysis of soluble protein extracts by heat prefractionation. Proteomics 2012; 12:938-43; PMID:22522800; http://dx.doi.org/10.1002/pmic.201100475
114. Meyer DE, Chilkoti A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. Nat Biotechnol 1999; 17:1112-5; http://dx.doi.org/10.1038/15100
115. Borovskii GB, Stupnikova IV, Antipina AI, Vladimirova SV, Voinikov VK. Accumulation of dehydrin-like proteins in the mitochondria of cereals in response to cold, freezing, drought and ABA treatment. BMC Plant Biol 2002; 2:5; http://dx.doi.org/10.1186/1471-2229-2-5
116. Banki MR, Feng L, Wood DW. Simple bioseparations using self-cleaving elastin-like polypeptide tags. Nat Methods 2005; 2:659-61; PMID:16074986; http://dx.doi.org/10.1038/nmeth787.
117. Bellucci JJ, Amiram M, Bhattacharyya J, McCafferty D, Chilkoti A. Three-in-one chromatography-free purification, tag removal, and site-specific modification of recombinant fusion proteins using sortase A and elastin-like polypeptides. Angew Chem Int Ed Engl 2013; 52:3703-8; PMID:23424160; http://dx.doi.org/10.1002/anie.201208292
118. Schellenberger V, Wang C-W, Geething NC, Spink BJ, Campbell A, To W, Scholle MD, Yin Y, Yao Y, Bogin O, et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat Biotechnol 2009; 27:1186-90; PMID:19915550; http://dx.doi.org/10.1038/nbt.1588
119. Garay RP, El-Gewely R, Armstrong JK, Garratty G, Richette P. Antibodies against polyethylene glycol in healthy subjects and in patients treated with PEG-conjugated agents. Expert Opin Drug Deliv 2012; 9:1319-23; PMID:22931049; http://dx.doi.org/10.1517/17425247.2012.720969
120. Peng Z, Oldfield CJ, Xue B, Mizianty MJ, Dunker AK, Kurgan L, Uversky VN. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci 2013; •••:6-10; PMID:23942625
121. Schwille P. Bottom-up synthetic biology: engineering in a tinkerer's world. Science 2011; 333:1252-4; PMID:21885774; http://dx.doi.org/10.1126/science.1211701
122. Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 1999; 65:382-8; PMID:10506413; http://dx.doi.org/10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
123. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 2008; 3:373-82; PMID:18533659; http://dx.doi.org/10.1021/cb800025k
124. Li T, Ma S-Y, Tang X-C, Nie L-Y, Huang H. Production and characterization of highly purified recombinant thymosin beta 4 in Escherichia coli. Protein Expr Purif 2013; 90:90-5; PMID:23711379; http://dx.doi.org/10.1016/j.pep.2013.05.006
125. Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 1988; 67:31-40; PMID:3047011; http://dx.doi.org/10.1016/0378-1119(88)90005-4
126. Karlsson R, Jendeberg L, Nilsson B, Nilsson J, Nygren PA. Direct and competitive kinetic analysis of the interaction between human IgG1 and a one domain analogue of protein A. J Immunol Methods 1995; 183:43-9; PMID:7602138; http://dx.doi.org/10.1016/0022-1759(95)00030-E
127. Hipps DS, Packman LC, Allen MD, Fuller C, Sakaguchi K, Appella E, Perham RN. The peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase complex of Bacillus stearothermophilus: preparation and characterization of its binding to the dihydrolipoyl dehydrogenase component. Biochem J 1994; 297:137-43; PMID:8280091
128. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science 1994; 263:802-5; PMID:8303295; http://dx.doi.org/10.1126/science.8303295
129. Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol 2005; 23:102-7; PMID:15580262; http://dx.doi.org/10.1038/nbt1044
130. Gautier A, Juillerat A, Heinis C, Corrêa IR Jr., Kindermann M, Beaufils F, Johnsson K. An engineered protein tag for multiprotein labeling in living cells. Chem Biol 2008; 15:128-36; PMID:18291317; http://dx.doi.org/10.1016/j.chembiol.2008.01.007

131. Watanabe T, Ito Y, Yamada T, Hashimoto M, Sekine S, Tanaka H. The roles of the C-terminal domain and type III domains of chitinase A1 from Bacillus circulans WL-12 in chitin degradation. J Bacteriol 1994; 176:4465-72; PMID:8045877
132. Yang D, Peng M, Yang H, Yang Q, Xu J. Expression, purification and characterization of Gloydius shedaoensis venom gloshedobin as Hsp70 fusion protein in Pichia pastoris. Protein Expr Purif 2009; 66:138-42; PMID:19286459; http://dx.doi.org/10.1016/j.pep.2009.03.003
133. Scior A, Preissler S, Koch M, Deuerling E. Directed PCR-free engineering of highly repetitive DNA sequences. BMC Biotechnol 2011; 11:87; PMID:21943395; http://dx.doi.org/10.1186/1472-6750-11-87
134. Zhou P, Lugovskoy AA, Wagner G. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR 2001; 20:11-4; PMID:11430750; http://dx.doi.org/10.1023/A:1011258906244
135. Buscaglia CA, Alfonso J, Campetella O, Frasch AC. Tandem amino acid repeats from Trypanosoma cruzi shed antigens increase the half-life of proteins in blood. Blood 1999; 93:2025-32; PMID:10068675
136. Alvarez P, Buscaglia CA, Campetella O. Improving protein pharmacokinetics by genetic fusion to simple amino acid sequences. J Biol Chem 2004; 279:3375-81; PMID:14612434; http://dx.doi.org/10.1074/jbc.M311356200
137. Paz-Ares L, Ross H, O'Brien M, Riviere A, Gatzemeier U, Von Pawel J, Kaukel E, Freitag L, Digel W, Bischoff H, et al. Phase III trial comparing paclitaxel poliglumex vs docetaxel in the second-line treatment of non-small-cell lung cancer. Br J Cancer 2008; 98:1608-13; PMID:18475293; http://dx.doi.org/10.1038/sj.bjc.6604372
138. Li C, Yu DF, Newman RA, Cabral F, Stephens LC, Hunter N, Milas L, Wallace S. Complete regression of well-established tumors using a novel water-soluble poly(L-glutamic acid)-paclitaxel conjugate. Cancer Res 1998; 58:2404-9; PMID:9622081
139. Moktan S, Perkins E, Kratz F, Raucher D. Thermal targeting of an acid-sensitive doxorubicin conjugate of elastin-like polypeptide enhances the therapeutic efficacy compared with the parent compound in vivo. Mol Cancer Ther 2012; 11:1547-56; PMID:22532601; http://dx.doi.org/10.1158/1535-7163.MCT-11-0998
140. Liu W, Dreher MR, Furgeson DY, Peixoto KV, Yuan H, Zalutsky MR, Chilkoti A. Tumor accumulation, degradation and pharmacokinetics of elastin-like polypeptides in nude mice. J Control Release 2006; 116:170-8; PMID:16919353; http://dx.doi.org/10.1016/j.jconrel.2006.06.026
141. Silva CIF, Skrzeszewska PJ, Golinska MD, Werten MWT, Eggink G, de Wolf FA. Tuning of collagen triple-helix stability in recombinant telechelic polymers. Biomacromolecules 2012; 13:1250-8; PMID:22404251; http://dx.doi.org/10.1021/bm300323q
142. Elliott S, Lorenzini T, Asher S, Aoki K, Brankow D, Buck L, Busse L, Chang D, Fuller J, Grant J, et al. Enhancement of therapeutic protein in vivo activities through glycoengineering. Nat Biotechnol 2003; 21:414-21; PMID:12612588; http://dx.doi.org/10.1038/nbt799
143. Zhang Y-B, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expr Purif 2004; 36:207-16; PMID:15249042; http://dx.doi.org/10.1016/j.pep.2004.04.020
144. Trabbic-Carlson K, Meyer DE, Liu L, Piervincenzi R, Nath N, LaBean T, Chilkoti A. Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: a role for surface hydrophobicity? Protein Eng Des Sel 2004; 17:57-66; PMID:14985538; http://dx.doi.org/10.1093/protein/gzh006
145. Trabbic-Carlson K, Liu L, Kim B, Chilkoti A. Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. Protein Sci 2004; 13:3274-84; PMID:15557268; http://dx.doi.org/10.1110/ps.04931604
146. Kolodziej KE, Pourfarzad F, de Boer E, Krpic S, Grosveld F, Strouboulis J. Optimal use of tandem biotin and V5 tags in ChIP assays. BMC Mol Biol 2009; 10:6; PMID:19196479; http://dx.doi.org/10.1186/1471-2199-10-6
147. Craven RA, Griffiths DJ, Sheldrick KS, Randall RE, Hagan IM, Carr AM. Vectors for the expression of tagged proteins in Schizosaccharomyces pombe. Gene 1998; 221:59-68; PMID:9852950; http://dx.doi.org/10.1016/S0378-1119(98)00434-X
148. Kochanczyk T, Jakimowicz P, Kreel A. Femtomolar Zn(II) affinity of minimal zinc hook peptides--a promising small tag for protein engineering. Chem Commun (Camb) 2013; 49:1312-4; PMID:23303248; http://dx.doi.org/10.1039/c2cc38174e
149. Sutton RB, Fasshauer D, Jahn R, Brunger AT. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature 1998; 395:347-53; PMID:9759724; http://dx.doi.org/10.1038/26412
150. Crooks RO, Rao T, Mason JM. Truncation, randomization, and selection: generation of a reduced length c-Jun antagonist that retains high interaction stability. J Biol Chem 2011; 286:29470-9; PMID:21697091; http://dx.doi.org/10.1074/jbc.M111.221267
151. Nock S, Spudich JA, Wagner P. Reversible, site-specific immobilization of polyarginine-tagged fusion proteins on mica surfaces. FEBS Lett 1997; 414:233-8; PMID:9315692; http://dx.doi.org/10.1016/S0014-5793(97)01040-5
152. Hoffmann C, Gaietta G, Bünemann M, Adams SR, Oberdorff-Maass S, Behr B, Vilardaga JP, Tsien RY, Ellisman MH, Lohse MJ. A FlAsH-based FRET approach to determine G protein-coupled receptor activation in living cells. Nat Methods 2005; 2:171-6; PMID:15782185; http://dx.doi.org/10.1038/nmeth742
153. Dekker C, Roe SM, McCormack EA, Beuron F, Pearl LH, Willison KR. The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. EMBO J 2011; 30:3078-90; PMID:21701561; http://dx.doi.org/10.1038/emboj.2011.208
154. Kim Y, Ho SO, Gassman NR, Korlann Y, Landorf EV, Collart FR, Weiss S. Efficient site-specific labeling of proteins via cysteines. Bioconjug Chem 2008; 19:786-91; PMID:18275130; http://dx.doi.org/10.1021/bc7002499