BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 8 2007, pages 950–956
doi:10.1093/bioinformatics/btm035

Structural bioinformatics

Local structural disorder imparts plasticity on linear motifs

Monika Fuxreiter, Peter Tompa and István Simon*

Institute of Enzymology, BRC, Hungarian Academy of Sciences, 1518 Budapest, P.O. Box 7, Hungary

Received on November 2, 2006; revised on January 11, 2007; accepted on January 27, 2007

Advance Access publication March 25, 2007

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: The dynamic nature of protein interaction networks requires fast and transient molecular switches. The underlying recognition motifs (linear motifs, LMs) are usually short and evolutionarily variable segments, which in several cases, such as phosphorylation sites or SH3-binding regions, fall into locally disordered regions. We probed the generality of this phenomenon by predicting the intrinsic disorder of all LM-containing proteins enlisted in the Eukaryotic Linear Motif (ELM) database.

Results: We demonstrated that LMs in average are embedded in locally unstructured regions, while their amino acid composition and charge/hydropathy properties exhibit a mixture characteristic of folded and disordered proteins. Overall, LMs are constructed by grafting a few specificity-determining residues favoring structural order on a highly flexible carrier region. These results establish a connection between LMs and molecular recognition elements of intrinsically unstructured proteins (IUPs), which realize a non-conventional mode of partner binding mostly in regulatory functions.

Contact: simon@enzim.hu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

A key level of the versatility of protein function at the cellular level derives from molecular adaptability in protein–protein interactions (PPIs). Communicating proteins are organized into graph-like networks denoted as interactome that offers a holistic description of the living cell (Gandhi *et al*., 2006; Gavin *et al*., 2006). At the structural level, PPIs can be realized by two distinct mechanisms (Neduva *et al*., 2005; Puntervoll *et al*., 2003). Traditionally, PPIs are associated with globular protein domains with a well-defined three-dimensional architecture. Such conserved functional units are specifically evolved to form high-affinity complexes with their partners. Recently, however, a good deal of the molecular recognition of proteins has been attributed to short segments rather than whole domains, first unraveled for the partners of the SH3 domain (Ren *et al*., 1993). A growing number of examples provide evidence that interaction via such linear motifs (LMs) offers an alternative way for protein-protein communication that can be beneficial in various cellular functions.

These short continuous regions with a few specificity determinants are hardly constrained in sequence, and thus could be easily switched on and off during evolution. Recognition through LMs results in interactions with micromolar affinities that underlie transient, reversible complexes, adapted for effective control. In light of these, experimentally described LMs are primarily found in signaling pathways, where they serve as consensus sites of post-translational modification or recognition elements in transient complexes (Neduva *et al*., 2005; Puntervoll *et al*., 2003). In general, these short segments are characterized by local flexibility. In fact, many of them, such as phosphorylation sites (Iakoucheva *et al*., 2004), deacetylation sites (Khan *et al*., 2005), SH3 interaction motifs (Beltrao *et al*., 2005), calmodulin binding sites (Radivojac *et al*., 2006), or recognition elements of 14-3-3 proteins (Bustos *et al*., 2006), have been found in locally disordered regions of their parent proteins.

The rationale of the connection between LMs and intrinsically unstructured/disordered proteins, often referred to as IUPs/IDPs, is that the latters are frequently associated with regulatory functions in signal transduction or transcription in eukaryotic proteomes (Dunker *et al*., 2002; Dyson *et al*., 2005; Liu *et al*., 2006; Minezaki *et al*., 2006; Tompa, 2002; Tompa, 2005; Tompa *et al*., 2006). These functions are carried out by molecular recognition that is linked to a binding-coupled folding process (Dyson *et al*., 2002; Oldfield *et al*., 2005; Tompa, 2005). Recently, it has been shown that partner recognition by unfolded segments is intimately related to the presence of short molecular recognition elements (MoREs) (Oldfield *et al*., 2005), a term in close analogy with preformed structural elements (Fuxreiter *et al*., 2004) or primary contact sites (Csizmok *et al*., 2005). Basically, the question addressed in this paper is the relationship and the range of possible correspondence between LMs and molecular recognition elements in IUPs.

To this end, we predicted disorder of LMs and their flanking regions for the experimentally characterized examples from the Eukaryotic Linear Motif (ELM) database (http://elm.eu.org) (Puntervoll *et al*., 2003), which has never been comprehensively tested in this regard. Possible disorder predictors range from neural network algorithms (PONDR, Peng *et al*., 2006) through amino-acid propensity based approaches (DisEMBL, Globplot, Linding *et al*., 2003a, Linding *et al*., 2003b) to those relying on direct physical principles of protein structure (IUPred, Dosztanyi *et al*., 2005a). Of these, we have selected for characterizing LMs two independent bioinformatics predictors, IUPred (Dosztanyi *et al*., 2005a; Dosztanyi *et al*., 2005b) and PONDR VSL1 (Obradovic *et al*., 2005; Peng *et al*., 2006),

*To whom correspondence should be addressed.

as well as the charge–hydropathy plot (Uversky, 2000) and the analysis of amino-acid compositions (Dunker *et al.*, 2001). These all show that LMs and their flanking regions are preferentially unstructured, resembling IUPs in many aspects. Within LMs, however, the uniformity breaks down as the IUP-like attributes mostly apply to the variable residues (non-restricted sites, NRSs), whereas specificity-determinant conserved (consensus) residues (restricted sites, RSs) take on the character of ordered, globular proteins. In general, LMs appear to have been constructed by grafting residues prone to be involved in protein-protein interactions onto a principally unstructured carrier sequence. Whereas this design of embedding appears to hold for the majority of LMs studied here, we must not overlook that there are exceptions mostly occuring in proteins associated with non-regulatory functions, where both the LM and the neighboring environment can adopt a well-defined fold.

Our results suggest that LMs and molecular recognition elements of IUPs significantly overlap, and in the majority of cases represent an equivalent set of motifs in modular proteins. The two definitions correspond to each other such as for phosphorylation sites or calmodulin-binding regions, and is expected to be so for more regulatory sites. In the case of association of metabolic proteins however, they likely conform to different concepts.

2 METHODS

2.1 Datasets

Out of the total of 116 different LMs available in the Eukaryotic Linear Motif (ELM) database (http://elm.eu.org), 72 with experimentally available instances were selected. These included 771 individual motifs in 549 proteins, sequences of which were extracted from the UNIPROT database and subjected to a 75% homology filtering. The resulting 469 sequences including 683 individual motifs were collected into a database. Sequence-inconsistent definitions detected in a few cases, e.g. LMs stretching over the total sequence length, were corrected. As a control, structural disorder was also calculated for proteins, for which independent biophysical evidence for disorder exists. These proteins can be downloaded form the DisProt database (http://www.disprot.org) (Sickmeier *et al.*, 2007). Amino acid frequencies in various subsets of the databases were also compared to those in globular proteins, as compiled in (Tompa, 2002).

2.2 Disorder prediction

Intrinsic disorder of LMs and their flanking regions was predicted by two independent predictors: the predictor of natural disordered regions, PONDR-VSL1 (http://www.pondr.com), and IUPred (http://iupred.enzim.hu) (Dosztanyi *et al.*, 2005a; Dosztanyi *et al.*, 2005b) algorithms. To bring the predictions on equal ground in terms of false positive prediction rate, PONDR-VSL1 scores were shifted by 0.15 downwards when the results were compared.

Since the lengths of LMs vary between 3 and 15 residues, the residue-based disorder values of each individual LM example were projected onto a uniform scale defined by the average LM length of six residues. The original position of all intermediate residues of LMs (with the exception of the first and last residue) were mapped onto a new scale by multiplying by a factor of $\{[5/(k-1)(i-1)]+1\}$, where k is the original length of the LM and i is the internal position of the given residue. The disorder values on the projected scale were obtained by averaging values falling into that given bin. In case of LMs shorter than six residues, the relevant bin values were duplicated. The difference between the disorder values averaged over the projected and original data were added to the new values by distributing it uniformly, to ensure that the average disorder of the LM remains the same.

2.3 Secondary structure and domain prediction

Secondary structure elements for all LM instances were predicted using the PROF algorithm (Ouali and King, 2000). Fraction of LM instances within ordered (helix and extended) secondary structure elements were obtained by averaging their frequences within each LM. Occurrence of LMs in domains were computed by the hmmpfam algorithm using the PFAM seed alignment database (Finn *et al.*, 2006).

2.4 Other methods

The regions encompassing LMs were also analyzed by two other approaches, which allowed the separate characterization of the tendency for order/disorder of conserved and variable subsites. The charge–hydropathy analysis relies on plotting mean net charge versus mean hydrophobicity of a sequence, as IUPs tend to have high net charge and low mean hydrophobicity, and are separated from globular proteins accordingly by a border line defined by the equation $\langle R\rangle = 2.785\langle H\rangle - 1.151$ (Uversky, 2000). The other approach is to compare the frequency of amino acids with IUPs and globular proteins, as IUPs tend to be enriched in P, E, K, S, G and Q, so-called disorder-promoting, and depleted in W, Y, F, C, I, L and V, so-called order-promoting, amino acids (Dunker *et al.*, 2001).

3 RESULTS

3.1 Average disorder profiles

Residue-based disorder was predicted for all LM-containing protein sequences by IUPred and PONDR VSL1. The disorder score within LMs was projected onto a uniform length, corresponding to the average of 6 residues (described in section 'Methods'). Furthermore, since different LMs include varying numbers of individual examples within the ELM database, the results were first averaged within each LM, and then over the 72 LMs to avoid over-representation of any LM (e.g. LIG-SH3). The resulting profiles are displayed in Figure 1. As explained in section 'Methods', values obtained by PONDR VSL1 were shifted by 0.15 downwards, so that predictions corresponded to the same false positive rate as for IUPred. In this way the two predictors provide the consensus result that the average score for LMs and their approximately 20 residue-long flanking segments on both sides is above 0.5, i.e. these regions can be considered as intrinsically disordered. Furthermore, an increased flexibility for approximately 100 residue-long flanking segments, can also be observed. If the threshold of 0.4, the average disorder of all unstructured protein segments in the DisProt database (Vucetic *et al.*, 2005) is used as a criterion, even these longer, 200 residue-long regions can be defined as unstructured. Overall, LMs and their embedding nearby environments exhibit excessive flexibility that is comparable to IUPs. Intriguingly, flanking segments on the two sides of LMs exhibit some asymmetry, the reason of which is not apparent at the present: the flanking disordered region is significantly longer on the C-terminal side, where disorder score is also about 5% higher.

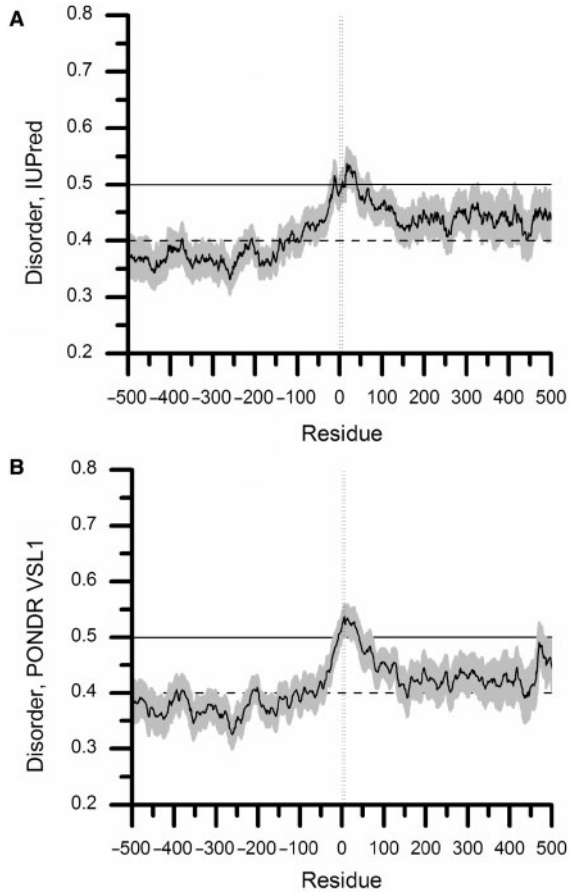

Fig. 1. Disorder profiles of proteins containing LMs. Disorder profiles by (**A**) IUPred and (**B**) PONDR VSL1 were computed in a 1000 residue frame around the LM motif. LMs of different lengths were projected to the same length, to formally fall between residues 1–6 (marked with vertical dotted lines). The predicted disorder values obtained by PONDR were shifted by 0.15 to correspond to the same false positive rate as those by IUPred (see Methods). A thin horizontal line at 0.5 shows the value above which the region is considered intrinsically unstructured, whereas a similar dashed line at 0.4 shows the average score for unstructured protein segments in DisProt (Sickmeier *et al.*, 2007). Values were first averaged within each LM, and then over the 72 LMs. Standard error of the mean values are displayed in light grey.

An additional noteworthy feature is that while the profile computed by the PONDR VSL1 method displays a single peak settled at the position of the LM, IUPred predicts a local downward spike at the same position. This resembles the disorder pattern of alpha MoREs (Oldfield *et al.*, 2005), where the recognition element is associated with slightly depressed disorder values compared to its unstructured environment. Such irregularity in the disorder pattern was suggested to indicate a transient structural element that can facilitate binding. This appears to be a common scenario for signaling, such as in the well-characterized examples of binding of p53 to MDM2, 4EBP1 to eIF4E and calcineurin to calmodulin (Oldfield *et al.*, 2005). This local feature disappears if the sequence of LMs and their 50-residue long flanking segments is randomized, which leaves the average disorder almost unchanged (data not shown). This suggests that local tendency for ordering of LMs is specific and sequence-dependent, whereas the intrinsic disorder of the flanking segments is inherent in their amino acid composition.

3.2 Amino acid composition

Amino acid frequencies of 683 individual LM examples and their 20-residue long flanking segments were analyzed and compared to those of globular proteins and IUPs in Figure 2A. The composition of LMs, as well as their 20-residue flanking segments shows significant deviations from the composition of globular proteins. Flanking regions are fairly similar to IUPs: they are enriched in flexible (disorder-promoting) amino acids, whereas depleted in hydrophobic and rigid (order-promoting) residues. On the other hand, they have considerably less charged residues than unstructured proteins, while they exhibit a larger excess of a small, polar residue, serine. LMs themselves have a unique amino acid composition, dissimilar to either globular or unstructured proteins, being abundant in both floppy and rigid residues, while depleted in amino acids of medium flexibility. LMs are enriched in hydrophobic residues W, L, F and Y as well as in charged residues R, and D, likely to play distinguished role in partner recognition. They are conspicuously depleted in G and A, possibly to simultaneously curb flexibility (G) and the overt tendency to form secondary structural elements (A), while preserving an open but structurally somewhat constrained state. The prevalence of proline in both LMs and their flanking segments underscores this logic due to its preference for an extended secondary structure motif, the polyproline-II helix.

The sequence of LMs can be dissected, as typical motifs are defined by a short sequence pattern (cf. Table S1 in Supplementary material), in which certain residues are of restricted identity (e.g. P in P..P for the SH3 binding sites), whereas others are of unrestricted identity (marked with '.' in the pattern). The first set of residues, which are either fully conserved or can vary to some extent, serve as specificity determinants of the LM, while the second set corresponds to fully variable positions that likely act as spacers. We reasoned that amino acid propensities of the two groups should deviate considerably if they serve disparate tasks in recognition. To probe this hypothesis, we computed the amino acid frequencies for 'RS' and 'NRS' positions separately (Fig. 2B). As anticipated, the compositions of these sites are markedly different, which is responsible for the aforementioned peculiarity of LM propensities. At the conserved positions either hydrophobic and rigid, or charged and flexible residues are preferred, as above, while in NRSs excessive flexibility, practically mirroring that of IUPs, can be observed. Of special functional status are prolines, which are in excess in RS positions and LM flanking regions as well, indicating their dual role as contact residues within LMs and promoters of an open structure outside LMs. Overall, the amino acid composition reveals the mixed nature of LMs and their embedding environments as extended disordered regions (flanking + NRSs), onto which a few specificity-determinant residues (RSs) have been grafted.

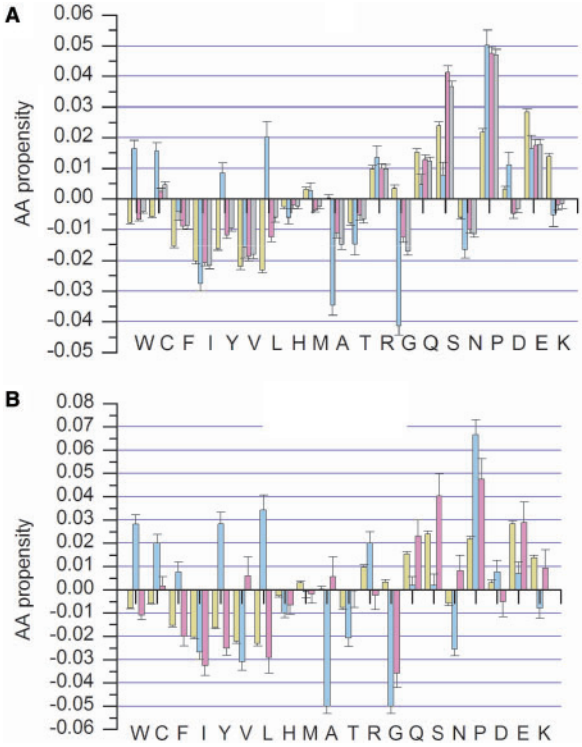

Fig. 2. Amino acid propensities of LMs and their flanking regions. Amino acid propensities of each data set are presented as the difference between its actual amino acid composition and that of globular proteins (Tompa, 2002): (**A**) IUPs of the DisProt database (yellow), LMs (cyan), 20 residue-long LM flanking segments (magenta), and LMs + 20 residue long flanking segments (light grey); (**B**) IUPs of the DisProt database (yellow), RSs (LM specificity determinant residues, cyan), NRSs (LM variable residues, magenta). In accord with the practice in the IUP field, frequencies in reference to those of globular domains are presented in the order of the Vihinen flexibility scale (Dunker *et al.*, 2001).

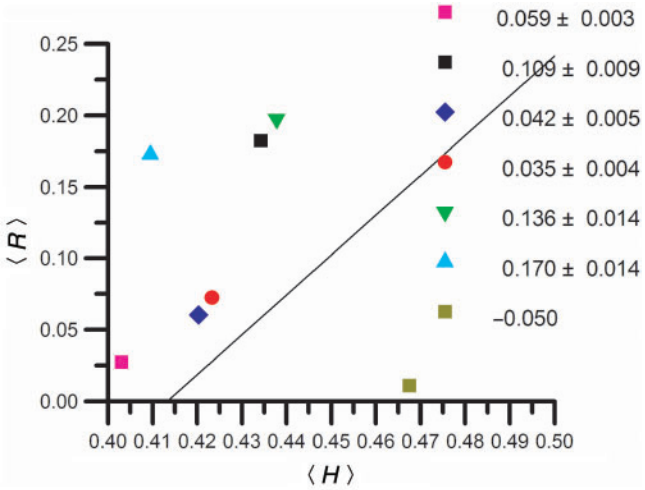

Fig. 3. Charge/hydropathy plots of LM subsets. Mean net charge ($\langle R \rangle$) and mean hydrophobicity ($\langle H \rangle$) is plotted for LMs (black square), their 20 residue-long flanking regions (blue diamonds), LMs + their flanking regions (red circles), RSs (specificity determinants of LMs, green down triangle), NRSs (variable residues of LMs, cyan up triangle), globular proteins (light brown squares) and IUPs (magenta squares). The grey line represents the borderline between IUPs and globular proteins as defined by Uversky (Uversky, 2000) $\langle R \rangle = 2.785\ \langle H \rangle - 1.151$. Legends express the distance from the Uversky line (in arbitrary units) with standard deviations.

3.3 Charge/hydropathy properties

One of the characteristic features that distinguish IUPs from globular proteins is their high mean net charge and low mean hydrophobicity (Uversky, 2000). As noted above, these parameters may take on unusual values within the LM region. To position LMs and their flanking segments in the globular-disordered plane thus defined, we computed their charge/hydropathy properties and averaged over all LMs (Fig. 3). Based on these features, both LMs and their nearby flanking segments lie in the disordered region of the Uversky plot. Interestingly, a level of net charge in LMs both in RS and NRS positions exceeding even that of IUPs, can be observed. These charge/hydrophathy ratios unveil the functional relevance of the unique amino acid composition of LMs: the high propensity of both hydrophobicity and charge at RS positions reflects their distinguished role as contacting residues in recognition, with interspersed variable positions selected more for flexibility/malleability. This pattern may also reveal an evolutionary compensatory mechanism, in which increased hydrophobicity brought about by residues involved in critical interactions has been compensated by an increased net charge of LM, preserving a significant tendency of disorder despite the presence of residues, which strongly favor order.

3.4 Functional classifications

The average disorder profiles suggest that the majority of LMs and their flanking regions are segments of intrinsic disorder within a more ordered environment. This observation is corroborated by the paucity of ordered secondary structure elements in LMs and their intermediate population in flanking segments (Fig. S1 and Table S2 in Supplementary material). Furthermore, LMs rarely appear in ordered PFAM domains, but if they do, the domain is almost invariably associated with post-translational modification, such as glycosylation, phosphorylation and fucosylation (Table S2 and unpublished data).

LMs represent a functionally very heterogeneous group, with proteins involved in either distinct post-translational modification reactions, or transient/permanent binding of other cellular constituents. In accord, individual analysis of LMs shows that a total of 56 out of 72 (78%) are locally disordered by both prediction methods using a 0.4 disorder threshold (cf. Fig. 1 and Table S1 in Supplementary material). When the distribution of the disorder of individual LMs is considered (Fig. 4), only 4% show a high level of disorder (0.8–1.0), the majority (74%) has an intermediate level (0.4–0.8), and only a smaller fraction (22%) tends to be ordered (0.0–0.4). High level of disorder however, is not due to the terminal location of the LM (Table S2). Certain functional categories, like transcription regulation (e.g. LIG_HP1_1) and protein-protein interactions

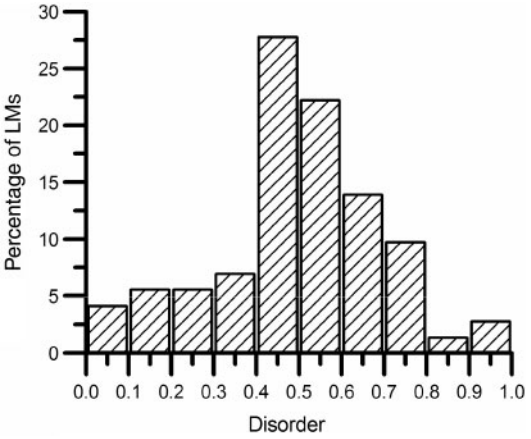

Fig. 4. The distribution of disorder in all LMs. The percentage of LMs with the given level of disorder, as predicted by the IUPred predictor, is plotted (cf. Table S1 in Supplementary material). 0.5 is the threshold above which the region is considered intrinsically unstructured, whereas 0.4 is the average score for unstructured protein segments in DisProt (Vucetic *et al.*, 2005).

(e.g. TRG_Golgi_diPhe_1) are overrepresented within those with a preference for local secondary structure. This observation suggests that protein-protein interactions in certain molecular settings tend to be promoted by pre-formed recognition elements, as already suggested (Fuxreiter *et al.*, 2004). Apparently, the induced folding in these examples is facilitated by the local structural preferences of the protein, in cases when rapid assembly of complexes is advantageous (e.g. transcription regulation). On the other hand, post-translational modification sites and many other PPI recognition elements are found to be associated with excessive flexibility, with a slight excess of the latter. Interestingly, proteins predicted to have less disordered LM regions (lower part of Table S1) do not necessarily have more predicted secondary structure (Table S2), which also attests to that the local effect of order-promoting hydrophobic RS residues is compensated in many cases by more disorder-promoting NRS side-chains. Thus, the special amino-acid composition of LMs, together with their flanking regions, limit local secondary structural preferences, thus creating a unique local environment specifically tailored for effective molecular recognition. Proteins with the least disordered LMs are mostly involved in membrane binding, participate in protein cargo or are targets for polysaccharide attachments. At a first and general approximation concerning biological functions, proteins involved in signal transduction, phosphorylation, cell proliferation (growth, adhesion) and endocytosis are primarily associated with the most unstructured LMs.

4 DISCUSSION

Dynamics of the cell is ensured by the information flow across precisely regulated protein–protein interaction networks. This process requires efficient communication between the participating proteins realized via their interactions. In classical terms, this is achieved via a tight, high-affinity fit of protein domains that possess stable three-dimensional structures. Evolution of such globular domains, however, is slower than that of short sequential motifs of a linear arrangement of specificity determinants. These latter, due to relying on a very few confined residues and limited structural constraints, may frequently arise and disappear by point mutations, and confer extreme adaptability to the interactome (Neduva *et al.*, 2005; Puntervoll *et al.*, 2003). Notwithstanding these long-term evolutionary benefits, key advantages of LMs reside in that they facilitate the communication between proteins by having less stringent requirements for association. The resulting low-affinity complexes are characterized by fast on and off binding rates that may serve regulatory purposes. Our interest was grasped by the fact that these very same features and functional advantages are repeatedly ascribed to IUPs, proteins or protein domains that function without a stable structure (Dunker *et al.*, 2002; Dyson *et al.*, 2005; Tompa, 2002; Tompa, 2005). The overlap, or possible mutual correspondence, of LMs and short recognition elements in IUPs is of significant appeal, and has been demonstrated in a handful of cases, such as phosphorylation sites (Iakoucheva *et al.*, 2004) and SH3 binding motifs (Beltrao *et al.*, 2005) (cf. also section 'Introduction'). The recent recognition that alternative splicing preferentially occurs in locally disordered regions of proteins, and removes short recognition elements (Romero *et al.*, 2006) extends the range of this possible correspondence. These examples raised the question if the molecular action of all LMs is based on the same principle of short recognition elements in unstructured regions.

To elaborate the relationship between these functional elements, we computed the disorder of each experimentally characterized non-homologous example of the ELM database (Puntervoll *et al.*, 2003) by two independent bioinformatic predictors and two propensity-based approaches. The methods consensually demonstrated that LMs and their 20-residue flanking segments have a markedly lower propensity to form ordered structure than the rest of the parent protein, with amino acid compositions comparable to that of IUPs enlisted in DisProt (Sickmeier *et al.*, 2007). LMs themselves, however, show an intriguingly complex pattern in that they are also enriched in hydrophobic residues as compared to their immediate environments, suggesting their distinguished role in recognition. Amino acid propensities as well as the hydropathy/charge values indicate a distinction between their conserved specificity determinants and variable residues: specificity determinant positions are abundant in hydrophobic and charged residues, and in many respects resemble globular proteins, whereas variable residues approximate features of IUPs. As a consequence, a local drop in plasticity is observed in the disorder profile computed by the IUPred method. In essence, the LM with its flanking region is a locally unstructured segment of the protein, onto which a couple of conserved, mostly hydrophobic, specificity-determinant residues are grafted.

Based on this characteristic structural pattern, we propose that LMs together with their flanking segments form a larger functional unit that can be considered as a super-secondary structural element. Such elements classically correspond to

structural/functional units within globular proteins, exemplified by the helix-turn-helix motif in DNA recognition (Luscombe *et al*., 2000) or the α-β-α Rossmann fold in nucleotide coenzyme binding (Rossmann *et al*., 1974). These elements are defined by a conserved spatial arrangement of secondary structural elements, maintained by a characteristic amino acid composition. The super-secondary element encompassing LMs, as defined here, conforms to this definition, as it consists of a functional unit composed of three structurally defined parts: a short, locally structured segment sandwiched between two disordered regions. As shown, the properties of the unstructured stretches are determined solely by their amino acid composition, while the actual primary sequence is critical only for the central part. This is not only an evolutionary economical solution to modulate protein-protein interactions, but also provides various benefits in the recognition process itself.

The ensuing binding mode, where determinant residues serve as contact sites, while variable residues within the LM as well as in its flanking environment ensures their proper orientation, has been often implicated in the interactions of IUPs with their partners, in the form of short recognition elements, described in various contexts as preformed structural elements (Fuxreiter *et al*., 2004), primary contact sites (Csizmok *et al*., 2005) or molecular recognition elements (Oldfield *et al*., 2005). Besides a proper control on the thermodynamics of binding, this binding mode also enables an increased speed of recognition (Dunker *et al*., 2002; Tompa, 2002), also formulated in mechanistic models such as protein fishing (Evans *et al*., 2002) or fly-casting (Shoemaker *et al*., 2000). These considerations argue that the molecular action of LMs is intertwined with the structural malleability of their embedding environment. Flanking segments enable the effective recognition of the target site that results in low-affinity contacts, which are of high specificity but readily reversible. Such modes of interactions are of significant benefit in regulation, signaling and network organization functions, as amply discussed in recent papers (Dunker *et al*., 2005; Haynes *et al*., 2006; Uversky *et al*., 2005). On the practical side, the recognition of this relationship also results in an improved predictability of such short motifs, as already demonstrated in the case of phosphorylation sites (Iakoucheva *et al*., 2004) and SH3 binding motifs (Beltrao *et al*., 2005).

According to this scenario, the molecular mode of action of LMs and recognition elements of IUPs overlap, and probably reflect a similar phenomenon viewed from different angles thus far. Whereas LMs have been identified as short sequence motifs critical in recognition functions, primary contact sites (Csizmok *et al*., 2005), preformed structural elements (Fuxreiter *et al*., 2004) and molecular recognition elments/features (Mohan *et al*., 2006; Oldfield *et al*., 2005) have been approached from the direction of protein disorder, emphasizing recognition segments embedded in such regions. Unity of these concepts is underlined by the key role of a few specificity determinants in all these recognition elements, and also by their local preferences for undergoing disorder-to-order transition upon molecular recognition (Dyson *et al*., 2002), which might already be manifested prior to binding (Fuxreiter *et al*., 2004). Apart from a minority of the cases when LMs are constituted of segments of ordered domains, these different concepts of short recognition elements express the same underlying physical and functional principles that provide a probably widespread solution to the dynamic control of protein-protein interaction networks (data not shown).

ACKNOWLEDGEMENTS

This work was funded by grants T049073, K60694 and F046164 from the Hungarian Science Fund (OTKA) as well as by Economic Competitiveness Operative Programme GVOP-3.2.1.-2004-04-0195/3.0. P.T. acknowledges support of the Wellcome Trust International Senior Research Fellowship GR067595 and M.F. for MRTN-CT-2005-019566 of the European FP6. The Bolyai János fellowship for M.F and Öveges grant from National Office for Research and Technology (I.S.) is also gratefully acknowledged. Funding to pay the Open Access publication charges was provided by the F046164 and T049073 grants of the Hungarian Scientific Research Fund (OTKA).

Conflict of Interest: none declared.

REFERENCES

Beltrao,P. *et al.* (2005) Comparative genomics and disorder prediction identify biologically relevant SH3 protein interactions. *PLoS Comput. Biol.*, **1**, 202–211.

Bustos,D.M. *et al.* (2006) Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins: Struct. Funct. Bioinform.*, **63**, 35–42.

Csizmok,V. *et al.* (2005) Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry*, **44**, 3955–3964.

Dosztanyi,Z. *et al.* (2005a) The pairwise energy content estimated from amino acid composition discriminates between folded and instrinsically unstructured proteins. *J. Mol. Biol.*, **347**, 827–839.

Dosztanyi,Z. *et al.* (2005b) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.

Dunker,A.K. *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graphics Modelling*, **19**, 26–59.

Dunker,A.K. *et al.* (2002) Intrinsic disorder and protein function. *Biochemistry*, **41**, 6573–6582.

Dunker,A.K. *et al.* (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, **272**, 5129–5148.

Dyson,H.J. *et al.* (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, **12**, 54–60.

Dyson,H.J. *et al.* (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell. Biol.*, **6**, 197–208.

Evans,P.R. *et al.* (2002) Endocytosis and vesicle trafficking. *Curr. Opin. Struct. Biol.*, **12**, 814–821.

Finn,R.D. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, D247–D251.

Fuxreiter,M. *et al.* (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.*, **338**, 1015–1026.

Gandhi,T.K. *et al.* (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat. Genet.*, **38**, 285–293.

Gavin,A.C. *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*, **440**, 631–636.

Haynes,C. *et al.* (2006) Intrinsic disorder is a common feature of hub proteins from fou eukaryotic interactomes. *PloS Comput. Biol.*, DOI: 10.1371/journal.pcbi.0020100.eor.

Iakoucheva,L.M. *et al.* (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, **32**, 1037–1049.

Khan,A.N. *et al*. (2005) Unstructured conformations are a substrate requirement for the Sir2 family of NAD-dependent protein deacetylases. *J. Biol. Chem*., **280**, 36073–36078.

Linding,R. *et al*. (2003a) Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.

Linding,R. *et al*. (2003b) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*., **31**, 3701–3708.

Liu,J. *et al*. (2006) Intrinsic Disorder in Transcription Factors. *Biochemistry*, **45**, 6873–6888.

Luscombe,N.M. *et al*. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol*., **1**, Reviews001.

Minezaki,Y. *et al*. (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol*., **359**, 1137–1149.

Mohan,A. *et al*. (2006) Analysis of molecular recognition features (MoRF). *J. Mol. Biol*., **362**, 1043–1059.

Neduva,V. *et al*. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett*., **579**, 3342–3345.

Obradovic,Z. *et al*. (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins*, **61** (Suppl. 7), 176–182.

Oldfield,C.J. *et al*. (2005) Coupled Folding and Binding with alpha-Helix-Forming Molecular Recognition Elements. *Biochemistry*, **44**, 12454–12470.

Ouali,M. and King,R.D. (2000) Cascaded multiple classifiers for secondary structure prediction. *Prot. Sci*., **9**, 1162–1176.

Peng,K. *et al*. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, **7**, 208.

Puntervoll,P. *et al*. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res*., **31**, 3625–3630.

Radivojac,P. *et al*. (2006) Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins*, **63**, 398–410.

Ren,R. *et al*. (1993) Identification of a ten-amino acid proline-rich SH3 binding site. *Science*, **259**, 1157–1161.

Romero,P.R. *et al*. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA*, **103**, 8390–8395.

Rossmann,M.G. *et al*. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature*, **250**, 194–199.

Shoemaker,B.A. *et al*. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA*, **97**, 8868–8873.

Sickmeier,M. *et al*. (2007) DisProt: the database of disordered proteins. *Nucleic Acids Res*., **35**, D786–D793.

Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem Sci*., **27**, 527–533.

Tompa,P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett*., **579**, 3346–3354.

Tompa,P. *et al*. (2006) Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J. Proteome Res*., **5**, 1996–2000.

Uversky,V.N. *et al*. (2000) Why are ''natively unfolded'' proteins unstructured under physiologic conditions? *Proteins*, **41**, 415–427.

Uversky,V.N. *et al*. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit*., **18**, 343–384.