

Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics

Charlotte S. Sørensen^{a,b} and Magnus Kjaergaard^{a,b,c,d,1}

^aDepartment of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark, DK-8000; ^bThe Danish Research Institute for Translational Neuroscience (DANDRITE), Aarhus University, Aarhus, Denmark, DK-8000; ^cAarhus Institute of Advanced Studies (AIAS), Aarhus University, Aarhus, Denmark, DK-8000; and ^dCenter for Proteins in Memory - PROMEMO, Danish National Research Foundation, Aarhus, Denmark, DK-8000

Edited by Alan R. Fersht, University of Cambridge, Cambridge, United Kingdom, and approved October 10, 2019 (received for review March 20, 2019)

Many multidomain proteins contain disordered linkers that regulate interdomain contacts, and thus the effective concentrations that govern intramolecular reactions. Effective concentrations are rarely measured experimentally, and therefore little is known about how they relate to linker architecture. We have directly measured the effective concentrations enforced by disordered protein linkers using a fluorescent biosensor. We show that effective concentrations follow simple geometric models based on polymer physics, offering an indirect method to probe the structural properties of the linker. The compaction of the disordered linker depends not only on net charge, but also on the type of charged residues. In contrast to theoretical predictions, we found that polyampholyte linkers can contract to similar dimensions as globular proteins. Hydrophobicity has little effect in itself, but aromatic residues lead to strong compaction, likely through π -interactions. Finally, we find that the individual contributors to chain compaction are not additive. We thus demonstrate that direct measurement of effective concentrations can be used in systematic studies of the relationship between sequence and structure of intrinsically disordered proteins. A quantitative understanding of the relationship between effective concentration and linker sequence will be crucial for understanding disorder-based allosteric regulation in multidomain proteins.

effective concentration | intrinsically disordered protein | flexible linker | polymer physics | fluorescent biosensor

Protein interactions are tightly regulated. The specificity is not only determined by the protein structure, but also by which molecules a protein encounters. Molecular encounters are often enhanced by a direct physical connection between the interaction partners, which changes the protein–protein interaction from a bi- to a unimolecular process. Many proteins thus primarily function by connecting other proteins, for example, an enzyme to its substrate or weakly interacting proteins to each other (1–3). Furthermore, multidomain proteins often contain long linkers that play a similar role for intramolecular interactions (4) and modulate phase separation (5). A physical connection often increases encounter rates by several orders of magnitude, which shifts equilibrium position of binding reactions and the rates of biochemical reactions by a similar amount (6, 7). The encounter frequency of such linked reactions is concentration-independent. Instead, it depends on the architecture of the connection, and therefore linker properties directly affect protein function.

Interdomain linkers often belong to the family of intrinsically disordered proteins (IDPs) (8). IDPs do not fold into stable structures, but form an ensemble of interconverting structures that may contain a range of transient interactions. Sufficiently long disordered linkers allow the tethered domains to contact in any orientation, and therefore they represent a generic mechanism for fusing domains, which cannot be accomplished by rigid proteins. For example, p300 coactivators consist of an enzymatic domain coupled to protein–protein interaction domains via disordered linkers (4). The connection between molecules is not static, but can vary in time and space to regulate the functions

of the tethered domains. This variation may result in allosteric regulation via, e.g., phosphorylation or protein binding in the linkers. Allostery usually describes the thermodynamic interaction between 2 ligands that are not in direct contact. Usually, this allostery is transmitted through structural changes in folded domains, but may also occur via changes in the structural ensembles of IDPs (9, 10). Structural changes of linker regions can affect the function of the tethered domains, and linkers thus provide a generic mechanism whereby an event can affect distant parts of a protein. To understand such allosteric effects, it is crucial to study the role of linkers quantitatively.

The functions of linkers can be understood quantitatively in terms of effective concentrations. For an intramolecular reaction, the encounter rate between tethered domains equals the rate of the same untethered reaction at a given concentration (11). This concentration is known as the effective concentration. Formally, the effective concentration is defined as the ratio of the equilibrium constants for 2 equivalent binding reactions, where one occurs intra- and one intermolecularly (12). When the linker is sufficiently long to join the binding sites without strain, the effective concentration is independent of what is linked and solely a property of the linker (13, 14). Intriguingly, this suggests that effective concentrations can be measured in a convenient model system and extrapolated to other systems. Effective concentrations can be measured by competition experiments, where a free ligand displaces a tethered ligand (12). Such measurements

Significance

Many proteins contain disordered linkers that join different functional units. Such linkers are crucial to many multidomain proteins, but are rarely studied in their own right. Here we study how intramolecular binding reactions are controlled by disordered protein linkers via effective concentrations. We show that the effective concentrations follow geometric models developed to describe polymers. The physical properties encoded into the linker sequence control the compaction of the linker and thus the ensuing effective concentrations. The system developed here provides a strategy to systematically investigate the relationship between sequence and compaction of intrinsically disordered proteins. We show here that such relationships directly affect biochemical binding equilibria by controlling the effective concentration.

Author contributions: C.S.S. and M.K. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: All data for this paper have been deposited in Figshare, <https://doi.org/10.6084/m9.figshare.10029254>, and in the *SI Appendix*.

¹To whom correspondence may be addressed. Email: magnus@mbg.au.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1904813116/-DCSupplemental.

have mostly been used in efforts to optimize multivalent drugs through avidity (15, 16). In molecular biology, effective concentrations are rarely measured experimentally, but usually estimated theoretically from the volume in which the tethered ligand is free to diffuse (17–21). While such simple geometric models are commonly used, they have not been tested in complex biological systems.

The simplest linker is a fully disordered chain. Fully disordered linkers are an attractive model for understanding effective concentrations, as they are well-described by theories borrowed from polymer physics. When the sizes of disordered proteins are measured as, e.g., end-to-end distances or the radius of hydration or gyration, they scale with the chain length following a power law such as:

$$\text{Size}(R_H, R_g, D_{\text{end-to-end}}) \propto N^\nu,$$

where N is the number of residues and ν is a scaling exponent determined by chain compaction. Such scaling laws underpin theoretical calculations of effective concentrations, as the chain length defines the radius of the accessible volume. The scaling exponent for effective concentrations is thus usually assumed to be -3ν , although such models have not been verified experimentally. Prediction of effective concentrations thus depend on the scaling exponent ν . On average, IDPs have been found to have ν values from 0.51 to 0.58 (22–24), but the scaling exponents of disordered proteins varies from about 0.4 for disordered states of foldable proteins to about 0.72 for highly charged IDPs (25). For reference, globular proteins and rigid rods have scaling exponents of 0.33 and 1, respectively. The sequence–compaction relationship of IDPs has been studied by correlating chain size with variations in sequence. Net charge dominates chain compaction through intrachain repulsion (22, 25–27). Furthermore, compaction is weakly correlated to hydrophobicity and weakly anticorrelated to proline content (22). The literature depicts a complicated relationship between polyampholyte strength and compaction, as the overall effect of polyampholyte interactions can cause compaction or expansion (28, 29). The complexity arises due to the patterning of charged residues (29–31), which leads to attractive interactions between some parts of the chain and repulsive interactions with others.

Here we investigate how effective concentrations in multidomain proteins depend on linker architecture. We directly measure effective concentrations for many disordered linkers with systematic changes in the physical properties of the linker. Our fluorescent biosensor for measurement of the effective concentrations provides a way to probe sequence–compaction relationships in intrinsically disordered proteins and relating these to biochemical function.

Materials and Methods

Preparation of DNA Constructs. DNA constructs were obtained from GenScript by insertion of synthetic genes between the NdeI and BamHI sites of a pET15b vector and subcloning of new linkers using unique NheI and KpnI sites flanking the linkers. Full protein sequences are given in *SI Appendix, Supplementary Materials*.

Protein Expression and Purification. All fusion protein constructs were expressed in BL21(DE3) cells in 50 mL ZYM-5052 autoinduction medium (32) supplied with 100 $\mu\text{g}/\text{mL}$ ampicillin and shaking at 120 rpm. The temperature was kept at 37 $^\circ\text{C}$ for 3 h and thereafter decreased to 18 $^\circ\text{C}$. The cells were harvested by centrifugation (15 min, 6,000 $\times g$) after 40 to 48 h, when the cultures had changed color, indicating mature fluorescent proteins. Bacterial pellets were lysed using B-PER Bacterial Protein Extraction Kit (Thermo Scientific) according to the manufacturer's protocol, and the lysate was applied to gravity flow columns packed with nickel Sepharose. After washing with 20 mM NaH_2PO_4 , pH 7.4, 0.5 M NaCl, 20 mM imidazole, fusion proteins were eluted by increasing the imidazole concentration to 0.5 M. Fusion proteins were subsequently purified using Strep-Tactin XT Superflow columns (IBA)

according to the manufacturer's instructions and dialyzed overnight into Tris-buffered saline (TBS). The MBD2 peptide was expressed overnight in BL21(DE3) cells at 37 $^\circ\text{C}$ in ZYM-5052 autoinduction medium with 100 $\mu\text{g}/\text{mL}$ ampicillin with shaking at 120 rpm. The cells were resuspended in 20 mM NaH_2PO_4 , pH 7.4, 0.5 M NaCl, 20 mM imidazole and lysed by heating to 80 $^\circ\text{C}$ for 20 min (33), and debris pelleted by centrifugation (15 min, 14,000 $\times g$). The peptide was purified by nickel Sepharose as before except a stepwise elution up to 0.5 M imidazole was used before dialysis into TBS. It was critical to prepare and concentrate the MBD2 peptide freshly and store it on ice until use. Protein concentrations were measured using A_{280} .

Measurement of Effective Concentrations. A total of 0.1 μM of each fusion protein was titrated with the MBD2 peptide through 16 serial 2-fold dilutions in TBS. The starting concentration was in the range of 1.6 to 2 mM for WT MBD2 and 3.3 mM for the V227A MBD2. Samples were analyzed in triplicate in black 386-well plates with 1 g/L bovine serum albumin (BSA; Fisher Scientific) added to prevent sticking. The FRET measurements were performed in a SpectraMax I3 plate reader using donor excitation at 500 nm, and emission detected in 25-nm-wide bands centered at 535 and 600 nm. The titration data were analyzed by nonlinear fitting in MATLAB to the standard fitting equation for 1:1 binding reactions with K_d replaced by $C_{e,\text{app}}$:

$$\frac{I_A}{I_D + I_A} = E_1 - \frac{E_2}{2} \times \left((C_{\text{MBD2}} + C_{e,\text{app}} + P) - \sqrt{(C_{\text{MBD2}} + C_{e,\text{app}} + P)^2 - 4P \times C_{\text{MBD2}}} \right),$$

where E_1 and E_2 are the apparent FRET values in the open and closed states and P is the concentration of the fusion protein. For titration with the WT MBD2 peptide, this determines an “apparent effective concentration,” which was multiplied by the affinity ratio of the WT and V227A peptides to produce the true effective concentration. The correction factor was established to be 30 by titration of the fusion protein containing the GS₁₂₀ linker with the V227A MBD2 peptide. Polymer scaling parameters were extracted by a linear fit to $\log(C_e)$ vs. $\log N$ to allow an equal weighting of the errors in C_e .

Diffusion Measurements by Fluorescence Correlation Spectroscopy (FCS). FCS measurements were recorded on samples containing ~ 10 nM of the 40-residue linker variant of each linker composition in TBS at 23 $^\circ\text{C}$. The fluorescent proteins were excited using the 488-nm laser, and the donor dye emission was acquired for 5 min in 10- μs time bins using a home-built confocal single molecule FRET instrument described previously (34). Autocorrelation curves were generated using multiptau 0.3.3. Datasets with high molecular weight bursts were eliminated. The dimensions of the confocal volume were determined with Atto488 as reference (35). The autocorrelation curves were fitted globally using Igor Pro-6.37 to the equation describing free diffusion through a Gaussian volume with a triplet state contribution (36):

$$G(\tau) = \left(1 - T + T \times e^{-\frac{\tau}{\tau_{\text{trip}}}} \right) \times \frac{1}{N} \times \frac{1}{\left(1 - \frac{\tau}{\tau_d} \right)} \times \frac{1}{\sqrt{1 + \left(\frac{\tau}{z_d} \right)^2} \times \left(\frac{\tau}{r_d} \right)}.$$

The parameters describing the confocal volume (r_d , z_d) were determined from the reference sample, and the parameters describing the triplet state dynamics (T , τ_{trip}) were fitted as global parameters across all diffusion measurements.

Results

Reporter Design. We designed a fusion protein inspired by FRET biosensors (37) to measure effective concentrations for different linker architectures. An exchangeable linker joins 2 protein domains that form an interaction pair. These domains are flanked by the fluorescent proteins mClover3 and mRuby3, which form a FRET pair (38) (Fig. 14). When the intramolecular complex is formed, the fluorescent proteins are brought into close contact, resulting in efficient FRET. The effective concentration is measured by following the FRET efficiency in a titration (12), where a free ligand displaces the intramolecular interaction. The ideal interaction pair is small, has a known 3D structure, is easy to produce in *Escherichia coli*, and binds tightly to ensure full ring-closing. Furthermore, in the bound state, the N terminus of one protein should be close to the C terminus of the other and vice versa. This ensures a high FRET efficiency in the closed state and allows even a short linker to join the domains without

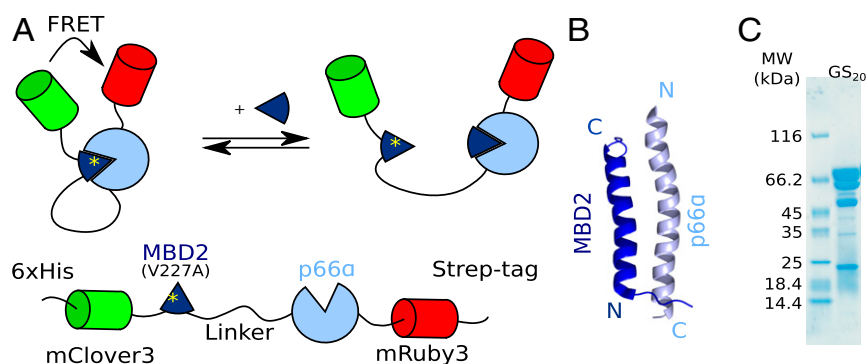


Fig. 1. A FRET biosensor for measuring effective concentrations. (A) Titration with free MBD2 peptide displaces the intramolecular interaction and results in a decrease in the FRET efficiency. (B) Structure of the interaction pair in the fusion protein: the antiparallel coiled coil formed between MBD2 and p66 α (PDB ID code 2L2L) (39). (C) SDS/PAGE gel of the purified fusion protein reveals internal cleavage sites in mRuby3. The cleavage products cannot be removed, but do not affect measurements.

strain. These constraints are ideally fulfilled by an antiparallel heterodimeric coiled coil such as the complex between MBD2 and p66 α used here (Fig. 1B) (39).

The fusion proteins have purification tags at both the N and C termini to allow parallel purification of many constructs by sequential affinity chromatography. After purification, SDS/PAGE revealed a major band corresponding to the expected molecular weight of the fusion protein (Fig. 1C). The 4 minor bands corresponded to proteolytic cleavage in the mRuby3 domain as revealed by mass spectrometry. These bands could not be removed by gel filtration. This suggests that the cleaved fluorescent proteins remain in a stable complex, although it is not clear whether it is fluorescent. A fraction of inactive fluorophores will decrease the FRET amplitude, but will not affect the midpoint of the titration and the measurements of effective concentrations.

Measurement of Effective Concentration. Fusion proteins were initially constructed with linkers consisting of (GS) $_n$ repeats ranging from 20 to 120 residues. A 240-residue GS-linker was also tested, but resulted in insoluble protein. Titration of all constructs with free WT MBD2 peptide resulted in a sigmoidal decrease of the proximity ratio (Fig. 2A), where the titration midpoint marks the apparent effective concentration. For increasing linker lengths, the titration midpoint was reached at lower peptide concentrations consistent with a decrease in effective concentration. Simultaneously, the proximity ratio of the posttitration baseline decreased with linker length consistent with a more expanded open form (Fig. 2A). Across the whole dataset, the pre- and posttransition proximity ratios varied unsystematically, which we believe was due to small differences in fluorophore maturation. This prevents determination of intramolecular distances

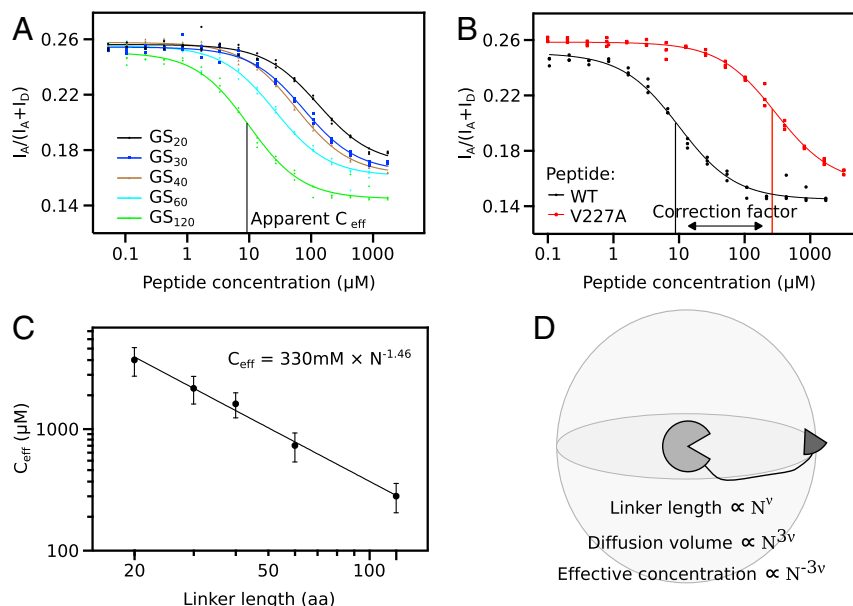


Fig. 2. Measurement of effective concentrations. (A) Titration curves of fusion proteins containing GS-linkers of variable length reveal that the effective concentration scales monotonously with linker length. (B) The correction factor corresponding to the difference in affinity between WT and V227A was determined by titration of fusion protein GS₁₂₀ containing the V227A mutation with both competitor peptides. (C) Effective concentrations follow a power law as revealed by the straight line in a log-log plot. The scaling exponent that is the focus of the remainder of the manuscript is derived from nonlinear fitting as indicated. (D) The expected geometric relation between scaling exponents for protein size, ν , and effective concentration. The linker defines a volume in which tethered ligand can diffuse, and the effective concentration is inversely proportional to this diffusion volume. Error bars represent 95% confidence interval.

the trends in the scaling coefficients are unlikely to be caused by aggregation or severe linker-induced collapse of the biosensor.

Are All Charged Residues Equal? To test if all charged residues affect linker compaction equally, we measured effective concentrations for linkers containing increasing amounts of the 4 charged residues. For each residue type, increased net charge per residue lead to chain expansion (Fig. 3C). Glutamate and aspartate caused an equal expansion: the scaling exponent changed gradually from -1.46 ($\nu = 0.49$) in the uncharged linker to ~ -2.1 ($\nu = \sim 0.7$) at a net charge per residue of 0.2. The scaling exponent did not increase further with increasing net charge, suggesting an upper limit that the linker does not expand beyond. This limit corresponds to the ν previously observed for highly charged IDPs (25). Lysine-containing linkers initially followed the expansion of negatively charged linkers, but continued up to a scaling exponent of ~ -2.4 ($\nu = \sim 0.8$). This is higher than the ν -values reported for any other disordered protein. In contrast, arginine-containing linkers have the opposite effect: residue fractions up to 0.1 had the same scaling exponent as GS-linkers. At higher fractions, the scaling exponent increased, but did not reach the same expansion as the other residue types even at the highest fractions permitted by solubility. The difference between arginine and lysine mirror their different roles in intermolecular interactions between disordered proteins (45). In total, the titration series demonstrated that charged residue types have a different impact on IDP compaction.

Polyampholyte Linkers. Most IDPs are polyampholytes, which entails that chain compaction is determined by the balance between attractive and repulsive interactions. Both simulation (30, 46) and experiments (46, 47) have shown that the patterning of charged residues have a large effect on polyampholyte compaction. To investigate the pure effect of ampholyte strength in the absence of charge patterning, we created linkers with a uniform distribution of equal numbers of positively and negatively charged residues. With the difference between lysine and arginine in mind, we also compared polyampholytes containing each of these 2 residues combined with glutamate. At low fractions of charged residues,

the scaling exponent was roughly constant for both polyampholyte series. However, there appeared to be a threshold after which the chain contracts dramatically for arginine-containing polyampholytes (Fig. 4A). The last data point at a fraction of charged residues of 0.67 suggested a scaling exponent of -1 ($\nu = 0.33$), which is the same compaction as a globular protein. This observation agrees qualitatively with the contraction caused by screening of charges in a polyampholyte (27), but contradicts the diagram-of-states representation of IDP classes (29, 30). In contrast, polyampholyte linkers containing lysine residues had approximately the same compaction for all fractions of charged residues.

The Effect of Proline Residues. Proline content is one of the main determinants of IDP compaction (22), in part because proline residues have less conformational freedom as they are restricted to certain ϕ/ψ -angles. Therefore, we gradually increased the fraction of proline residues in the linker (Fig. 4B). The scaling exponent increased to a plateau at ~ -1.7 ($\nu = \sim 0.57$), which is reached at a proline fraction of 0.1, which is similar to the dimensions of chemically denatured proteins (41). Proline thus expanded the linker to a lesser extent than charged residues, in agreement with previous studies (22). Unlike the previous series, the expansion with increasing proline content appeared non-monotonic. This could suggest that different effects dominate at different fractional content of proline, e.g., the propensity to form polyproline II helices, although it is difficult to draw a firm conclusion due to the experimental errors.

The Effect of Hydrophobic Interactions. It is unclear how hydrophobicity affects compaction of disordered proteins. In IDPs, hydrophobicity is weakly anticorrelated with ν (22). However, disordered states of foldable protein, which are much more hydrophobic than IDPs, form both compact (25) and expanded states (48). To assess the effect of hydrophobicity systematically, leucine residues were introduced into the linker. We chose leucine as it is among the most hydrophobic of the nonaromatic residues (49), but is not β -branched and thus less likely to perturb backbone dihedral distributions. Due to solubility, the linker sequence can only be extended up to leucine fractions of 0.2.

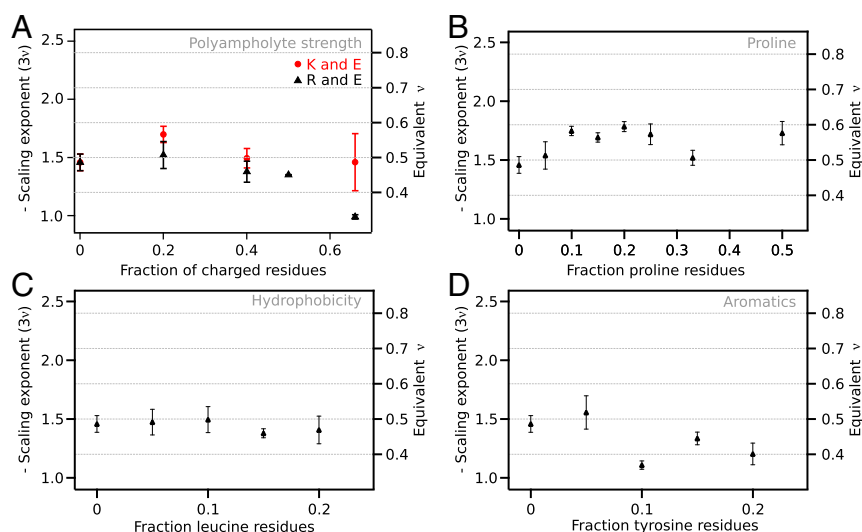


Fig. 4. The effect of polyampholyte strength and proline, leucine, and tyrosine residues on scaling exponents. (A) Scaling exponents from neutral polyampholyte linkers with an equal proportion of glutamate residues and either lysine or arginine. The scaling exponents show a strong compaction to a globular chain for polyampholytes containing arginine only. (B) Chain expansion shows a complex dependence on proline content. At low fractions of proline residues, proline residues lead to linker expansion, which may be reversed at high proline fractions. (C) Hydrophobicity was increased by introduction of leucine residues, but led to practically no change in chain compaction. (D) Tyrosine residues led to strong compaction, likely due to π -interactions. Error bars are SE estimated from the fit.

Introduction of leucine residues led to a small decrease in scaling exponent to -1.4 ($\nu = 0.47$; Fig. 4C). This qualitatively agreed with a weak anticorrelation between size and hydrophobicity.

π -Interactions between Aromatic Residues. Interactions between π -electrons in aromatic side chains have a strong potential to induce intrachain interaction, most recently demonstrated by their effect on liquid–liquid phase separation (45). We introduced tyrosine residues as it is the least hydrophobic aromatic amino acid and has the largest effect on phase separation (45, 50, 51). Tyrosine residues caused a noticeable reduction in the scaling exponents already at a fractional content of 0.1, where the scaling exponent was ~ -1.2 ($\nu = 0.4$; Fig. 4D). The contraction was smaller than the contraction observed for poly-ampholytes, although the fractional content of aromatic residues cannot be increased as far due to insolubility. As for proline, the contraction appears to be nonmonotonous, although it is difficult to state decisively due to experimental errors. However, the effect of tyrosine residues was larger than that of leucine, which suggests that π -interactions are more important to chain compaction than hydrophobicity.

Are Contributions to Chain Compaction Additive? Previous studies have identified individual factors that affect IDP compaction, but have not clarified whether they are additive. We therefore constructed linker series simultaneously increasing net charge and proline content. When probed alone, proline expanded linkers measurably already at a fraction of 0.05. However, when proline was introduced together with a charged residue, it did not lead to further expansion, but simply resulted in identical scaling exponents as glutamate-only linkers (Fig. 5A). This suggested that expansion caused by chain rigidity and net charge are not additive. Instead, the mixed sequences were simply dominated by the strongest individual effect.

Does Hydrophobicity Modify Charge Expansion? Hydrophobicity had a surprisingly small effect when investigated alone. To further test this conclusion in the context of additivity, we created mixed linker series containing glutamate and leucine residues in equal proportion. Again, the scaling exponent followed that of charge series alone. This series changed 2 parameters at once, so we designed a linker series where either the glutamate content or the leucine content was varied and the other held constant at a residue fraction of 0.1. When the charge expansion was probed in the context of a fraction of 0.1 leucine residues, the curve again followed the glutamate-only series (Fig. 5B). When the fraction of leucine residues was increased within the limits permitted by solubility, the scaling exponent decreases slightly, although within error (Fig. 5C). In total, these data suggest that hydrophobicity in itself has a vanishing effect on the compaction of IDPs, at least

within the limits of protein solubility. Furthermore, they suggest that the factors that affect IDP compaction are not necessarily additive.

Discussion

Linkers control many biochemical reactions via the effective concentration. Here we have shown that effective concentrations in multidomain proteins with disordered linkers follow polymer scaling laws. We have thus experimentally validated the geometric models commonly used to estimate effective concentrations, but also show that the effective concentration depends strongly on linker sequence. For a 100-residue linker, the difference between ν values of 0.4 and 0.7 corresponds to a 63-fold change in effective concentration assuming a constant prefactor. This can be the difference between an intramolecular interaction being saturated or hardly formed at all. Changes in the linkers following, e.g., ligand binding or posttranslational modification may thus be directly transmitted into allosteric regulation of the domains tethered at the end. Linkers may thus be one of the most direct examples of how the structural properties of intrinsically disordered proteins affect biochemical function, underscoring the need to understand the relationship between IDP sequence and compaction. In the short term, direct measurement of the effective concentrations using the system developed here may help us understand allostery in IDPs (10).

Sequence–Structure Relationships in IDPs. Polymer models have been used successfully to describe the structural properties of IDPs and are the foundation for theoretical predictions of effective concentrations. Here we show that the relationship can be reversed: measurement of effective concentrations is an efficient way to parametrize polymer descriptions of IDPs and describe the relationship between sequence and compaction. The range of scaling exponents observed here agrees well with the range of scaling exponents of disordered proteins described previously. This validates the assumption of a scaling coefficient for effective concentrations equal to -3ν , and bridges the present work to the existing literature on IDP compaction. While we study inter-domain linkers, it is likely that the conclusions can be generalized to other types of IDPs, although it is necessary to keep a number of caveats in mind. Fused domains may artificially affect the properties of a disordered segment, and it is thus important to exclude such effects. In our biosensors, the linker constitutes 3 to 17% of the protein, and its contributions are thus swamped in most measurements. Therefore, we cannot exclude interactions between the linker and the folded domains, but believe that the systematic changes with linker length and composition make contributions from spurious interactions unlikely.

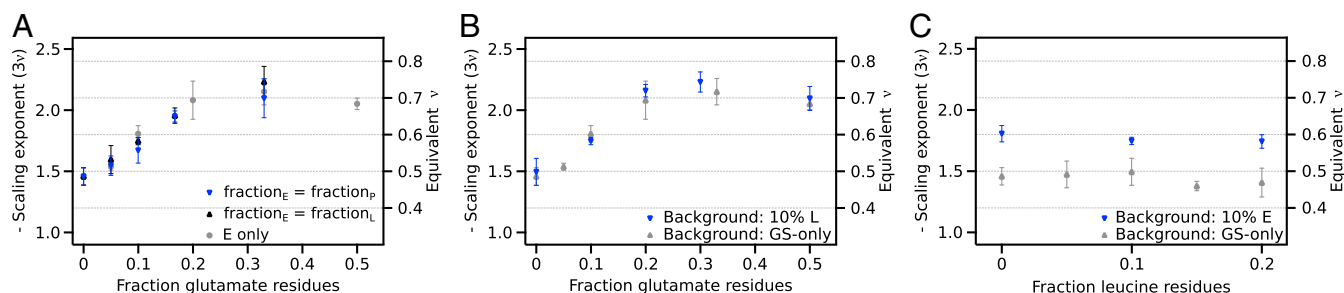


Fig. 5. Lack of additivity of individual contributions. (A) Proline or leucine residues were introduced in equal proportion to glutamate residues, but the scaling coefficient follows that of the glutamate-only series. This suggests that the expansion caused by net charge and chain stiffness are not additive. (B) Charge expansion caused by glutamate against a constant fraction of 0.1 of leucine residues shows that hydrophobicity has a negligible effect. (C) Increasing fractions of leucine only has a negligible effect on the dimensions of the linkers expanded by charge. Error bars represent SE.

Net Charge. Net charge is the strongest predictor of the compaction of IDPs (22, 25, 26). Previous studies suggest that the maximal expansion is reached at a net charge per residue of ~ 0.4 . In contrast, we find a value of ~ 0.2 (Fig. 3C). This difference may be due to sequence differences or a fundamental difference between a linker and an isolated chain. In previous studies, the charge expansion occurred against a complex background sequence with compensating attractive interactions. This can likely not be avoided, but is minimized in our linker series. Our value may thus represent a minimal estimate of how much charge density it takes to fully expand an otherwise inert IDP, whereas the value of around 0.4 may be more relevant for complex sequences with additional attractive interactions. Alternatively, linkers may be special because, e.g., the excluded volume of the attached domains favor expanded chains. Additionally, in the closed biosensor, the ends of the linkers are brought into close proximity, which may increase the linkers' sensitivity to charge repulsion compared to an isolated chain. Another key difference is the role of arginine residues. Arginine form attractive interactions that partially compensate for the charge–charge repulsion. This recapitulates the role of arginine in self-association of IDPs during liquid–liquid phase separation, and may thus be due to its capacity to form π -interactions (45, 51). Conveniently, for most positively charged protein sequences, this effect may be offset by the higher than average repulsion of lysine residues. Therefore, chain properties may be well-described in terms of the net charge density as long as lysine and arginine are equally common.

Polyampholyte Sequences. The compaction of polyampholytes is determined by the balance between attractive and repulsive interactions. As controlled experiments on polyampholytic IDPs have been scarce, organic polyampholytes have been used as models to understand the impact of ampholyte strength on IDP structure. Organic polyampholytes form a range of compact structures and, at high charge densities, they eventually become almost globular, albeit with a liquid-like internal structure (52). It is not clear how well such models describe proteins, especially because many proteins have a patchy distribution of charged residues. In one case, increase of the ionic strengths led to expansion of a polyampholytic protein (27), which suggested that polyampholyte interactions are overall attractive. This conclusion is also supported by the compact state adopted by the complex between 2 oppositely charged IDPs (53). On the contrary, a computational study of polyampholyte sequences suggested that the compact state only arises if charges are unevenly distributed, whereas a well-mixed polyampholyte was predicted to form expanded coils (30). This is summarized in the diagram-of-states description of IDPs, where increase of the strength of neutral polyampholyte leads to a globule-to-coil transition (29). Here we find that increase of the polyampholyte strength led to compaction (Fig. 4A), but only for polyampholytes containing arginine. This mirrors the difference we see between the positively charged residues probed in isolation. In the presence of the attractive π -interactions in arginine residues, even a perfectly mixed polyampholyte can contract, whereas the lysine-containing ampholytes remain relatively expanded as predicted previously (30). These polyampholytes resembles the mixed charge domains of nuclear speckle proteins, which contain repeats of positively and negatively charged residues. Increasing the fractions of arginine residues increases the propensity to condensate, which is consistent with the greater propensity for intramolecular self-association observed here (54). Intriguingly, this shows that the ratio between the 2 positively charged residues can tune the functional properties of IDPs *in vivo*.

Hydrophobic Side Chains in IDPs. Hydrophobic side chains in IDPs are mostly solvent-exposed. Contraction may bring such side chains in proximity to interact and thus form a partial protection

from the solvent. Therefore, it was expected that hydrophobicity would be anticorrelated with ν in disordered proteins (22, 25). Recently, this conclusion was questioned, as several unfolded, but foldable, proteins form relatively expanded chains ($\nu = 0.54$) despite their hydrophobicity (48). By testing the effect of hydrophobicity in a variety of sequence contexts, we found that increasing the hydrophobicity did not lead to a noticeable contraction of the linker. One explanation for this is that the disorder of the chain prevents the proteins from forming even partially desolvated hydrophobic interactions. In complex sequences, hydrophobicity correlates with other factors that could cause chain compaction. Such confounders could explain the correlation of hydrophobicity with the compaction of unfolded states of foldable proteins (25). A key candidate for such a confounder is aromatic residues, which we found to induce strong compaction of the linkers. A tyrosine fraction of 0.1 was thus sufficient to contract the linker to most compact dimensions observed for unfolded, but foldable, proteins (25). Similarly, tyrosine drives phase separation of IDPs to a much higher extent than leucine (55), providing the equivalent finding in intermolecular IDP interactions. We note that the most compact unfolded, but foldable, proteins in previous studies also have a higher fraction of aromatic residues. Aromatic residues thus potentially explain the disagreement (25, 49) on the effect of hydrophobicity in disordered proteins.

The Additivity of IDP Compaction. Several properties affect chain compaction, but how do they add up? Both charge and proline residues expand the linker, but the combination of the two did not expand the linker more than charge alone. This demonstrates that not all contributions to chain compaction are additive. A likely explanation is that rigidity added by proline residues can be accommodated inside the ensemble expanded by charge–charge repulsion. The reason net charge density is such a good predictor of the compaction of IDPs may thus be that the strongest effect dominates in the absence of additivity. The most likely explanation for the different threshold for charge expansion is thus the presence of compensating attractive interactions. This suggests an additivity between some factors, but likely not all, and demonstrates that much remains to be uncovered about sequence–compaction relationship of IDPs.

The Value of Synthetic IDPs. Our present understanding of the relationship between sequence and structure in IDPs is mainly based on the study of natural proteins. However, it is inherently difficult to draw general conclusions from a small set of examples. In contrast, we have studied synthetic linkers never seen in nature. The synthetic linkers allow tight control over the physical properties of the linker, which is crucial for hypothesis testing. Experiments on synthetic IDPs are thus a natural step for critically evaluating our understanding of IDPs. Synthetic DNA has removed the need for ingenious cloning strategies used previously (56), leaving protein preparation as the major bottleneck. For artificial proteins spanning a range of physical properties, this can, however, be a major challenge. We have, for example, not been able to make our linkers in isolation yet. The fusion protein used here serves both as solubility tags and a reporter system operating at nanomolar concentrations. These are likely the key factors that have allowed us to study a broader range of IDPs than previous studies. Although the potential of spurious interactions are intrinsic to such fusion proteins, they may provide useful tools for future investigations of sequence–structure relationships in IDPs. An exciting future direction would thus be a direct comparison of the scaling exponents determined from size and effective concentrations on the same sequences. Such measurements could help delineate in which ways the sequence–structure relationship of linkers differ from isolated IDPs.

Data Availability. All data discussed in the paper can be found in the [SI Appendix](#) and Figshare, doi:10.6084/m9.figshare.10029254.

ACKNOWLEDGMENTS. This work was supported by grants to M.K. from the “Young Investigator Program” of the Villum Foundation; the AIAS COFUND

program funded by the EU FP7 Cofund programme (Agreement no. 754513); and PROMEMO – Center for Proteins in Memory, a Center of Excellence funded by the Danish National Research Foundation (Grant Number DNRF133). We thank Birthe B. Kragelund, Mateusz Dyla, and Xavier Warner for critical comments to this manuscript; and Anna Marie Nielsen and Tanja Klymchuk for technical assistance.

1. L. K. Langeberg, J. D. Scott, Signalling scaffolds and local organization of cellular behaviour. *Nat. Rev. Mol. Cell Biol.* **16**, 232–244 (2015).
2. R. Nussinov, B. Ma, C. J. Tsai, A broad view of scaffolding suggests that scaffolding proteins can actively control regulation and signaling of multienzyme complexes through allostery. *Biochim. Biophys. Acta* **1834**, 820–829 (2013).
3. M. C. Good, J. G. Zalatan, W. A. Lim, Scaffold proteins: Hubs for controlling the flow of cellular information. *Science* **332**, 680–686 (2011).
4. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
5. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, 1–31 (2017).
6. T. Kitazawa *et al.*, A bispecific antibody to factors IXa and X restores factor VIII hemostatic activity in a hemophilia A model. *Nat. Med.* **18**, 1570–1574 (2012).
7. E. C. M. Greenwald, J. M. Redden, K. L. Dodge-Kafka, J. J. Saucerman, Scaffold state switching amplifies, accelerates, and insulates protein kinase C signaling. *J. Biol. Chem.* **289**, 2353–2360 (2014).
8. M. S. Cortese, V. N. Uversky, A. K. Dunker, Intrinsic disorder in scaffold proteins: Getting more from less. *Prog. Biophys. Mol. Biol.* **98**, 85–106 (2008).
9. E. Papaleo *et al.*, The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* **116**, 6391–6423 (2016).
10. P. Tompa, Multisteric regulation by structural disorder in modular signaling proteins: An extension of the concept of allostery. *Chem. Rev.* **114**, 6715–6732 (2014).
11. M. I. Page, W. P. Jencks, Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1678–1683 (1971).
12. V. M. Krishnamurthy, V. Semetey, P. J. Bracher, N. Shen, G. M. Whitesides, Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J. Am. Chem. Soc.* **129**, 1312–1320 (2007).
13. J. M. Gargano, T. Ngo, J. Y. Kim, D. W. K. Acheson, W. J. Lees, Multivalent inhibition of AB(5) toxins. *J. Am. Chem. Soc.* **123**, 12909–12910 (2001).
14. M. Li, H. Cao, L. Lai, Z. Liu, Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state? *Protein Sci.* **27**, 1600–1610 (2018).
15. E. T. Mack *et al.*, Dependence of avidity on linker length for a bivalent ligand-bivalent receptor model system. *J. Am. Chem. Soc.* **134**, 333–345 (2012).
16. H. X. Zhou, Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *J. Mol. Biol.* **329**, 1–8 (2003).
17. L. C. Timpe, L. Peller, A random flight chain model for the tether of the Shaker K⁺ channel inactivation domain. *Biophys. J.* **69**, 2415–2418 (1995).
18. D. J. Diestler, E. W. Knapp, Statistical mechanics of the stability of multivalent ligand-receptor complexes. *J. Phys. Chem. C* **114**, 5287–5304 (2010).
19. D. J. Diestler, E. W. Knapp, Statistical thermodynamics of the stability of multivalent ligand-receptor complexes. *Phys. Rev. Lett.* **100**, 178101 (2008).
20. W. Borchers *et al.*, Optimal affinity enhancement by a conserved flexible linker controls p53 mimicry in MdmX. *Biophys. J.* **112**, 2038–2042 (2017).
21. K. P. Sherry, S. E. Johnson, C. L. Hatem, A. Majumdar, D. Barrick, Effects of linker length and transient secondary structure elements in the intrinsically disordered notch RAM region on notch signaling. *J. Mol. Biol.* **427**, 3587–3597 (2015).
22. J. A. Marsh, J. D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).
23. P. Bernadó, D. I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167 (2012).
24. G. Fuertes *et al.*, Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6342–E6351 (2017).
25. H. Hofmann *et al.*, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16155–16160 (2012).
26. A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, R. V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8183–8188 (2010).
27. S. Müller-Spáth *et al.*, From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14609–14614 (2010).
28. B. Schuler, A. Soranno, H. Hofmann, D. Nettels, Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biophys.* **45**, 207–231 (2016).
29. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
30. R. K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13392–13397 (2013).
31. E. W. Martin *et al.*, Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J. Am. Chem. Soc.* **138**, 15323–15335 (2016).
32. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
33. C. Kalthoff, A novel strategy for the purification of recombinantly expressed unstructured protein domains. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **786**, 247–254 (2003).
34. M. Kjaergaard *et al.*, Oligomer diversity during the aggregation of the repeat region of Tau. *ACS Chem. Neurosci.* **9**, 3060–3071 (2018).
35. P. Kapusta, *Absolute Diffusion Coefficients: Compilation of Reference Data for FCS Calibration*. (Application note rev. 1, PicoQuant GmbH, Berlin, 2010).
36. P. Schwille, E. Haustein, Fluorescence correlation spectroscopy—An introduction to its concepts and applications. *Spectroscopy (Springf.)* **94**, (2001).
37. J. Zhang, Y. Ma, S. S. Taylor, R. Y. Tsien, Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14997–15002 (2001).
38. B. T. Bajar *et al.*, Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, 20889 (2016).
39. M. N. Gnanaprasam *et al.*, p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7487–7492 (2011).
40. A. S. Holehouse, K. Garai, N. Lyle, A. Vitalis, R. V. Pappu, Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *J. Am. Chem. Soc.* **137**, 2984–2995 (2015).
41. J. E. Kohn *et al.*, Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491–12496 (2004).
42. A. Mittal, A. S. Holehouse, M. C. Cohan, R. V. Pappu, Sequence-to-Conformation relationships of disordered regions tethered to folded domains of proteins. *J. Mol. Biol.* **430**, 2403–2421 (2018).
43. P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).
44. G. Tria, H. D. T. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217 (2015).
45. J. Wang *et al.*, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16 (2018).
46. R. K. Das, Y. Huang, A. H. Phillips, R. W. Kriwacki, R. V. Pappu, Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5616–5621 (2016).
47. K. P. Sherry, R. K. Das, R. V. Pappu, D. Barrick, Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9243–E9252 (2017).
48. J. A. Riback *et al.*, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).
49. W. C. Wimley, S. H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **3**, 842–848 (1996).
50. M. Kato *et al.*, Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
51. T. J. Nott *et al.*, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
52. A. V. Dobrynin, Theory and simulations of charged polymers: From solution properties to polymeric nanomaterials. *Curr. Opin. Colloid Interface Sci.* **13**, 376–388 (2008).
53. A. Borgia *et al.*, Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
54. J. A. Greig *et al.*, Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *bioRxiv*:10.1101/771592 (16 September 2019).
55. Y. Lin, S. L. Currie, M. K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120 (2017).
56. T. H. Evers, E. M. van Dongen, A. C. Faesen, E. W. Meijer, M. Merkx, Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry* **45**, 13183–13192 (2006).