THE PROTEIN SOCIETY

A polymetamorphic protein

Katie L. Stewart, Eric D. Dodds, Vicki H. Wysocki, and Matthew H. J. Cordes*

Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona

Received 30 November 2012; Revised 25 February 2013; Accepted 1 March 2013
DOI: 10.1002/pro.2248
Published online 8 March 2013 proteinscience.org

Abstract: Arc repressor is a homodimeric protein with a ribbon-helix–helix fold. A single polar-to-hydrophobic substitution (N11L) at a solvent-exposed position leads to population of an alternate dimeric fold in which 3_{10} helices replace a β-sheet. Here we find that the variant Q9V/N11L/R13V (S-VLV), with two additional polar-to-hydrophobic surface mutations in the same β-sheet, forms a highly stable, reversibly folded octamer with approximately half the α-helical content of wild-type Arc. At low protein concentration and low ionic strength, S-VLV also populates both dimeric topologies previously observed for N11L, as judged by NMR chemical shift comparisons. Thus, accumulation of simple hydrophobic mutations in Arc progressively reduces fold specificity, leading first to a sequence with two folds and then to a manifold bridge sequence with at least three different topologies. Residues 9–14 of S-VLV form a highly hydrophobic stretch that is predicted to be amyloidogenic, but we do not observe aggregates of higher order than octamer. Increases in sequence hydrophobicity can promote amyloid aggregation but also exert broader and more complex effects on fold specificity. Altered native folds, changes in fold coupled to oligomerization, toxic pre-amyloid oligomers, and amyloid fibrils may represent a near continuum of accessible alternatives in protein structure space.

Keywords: sequence-structure relationship; structural degeneracy; polar-to-hydrophobic substitution; structural evolution, fold switching

Introduction

The structural repertoire of proteins includes more than a thousand domain folds, but most individual natural protein sequences encode a single, specific native fold that is robust to mutation and conserved across long evolutionary distances. This high fold specificity and conservation gives the appearance that each protein is locked into a unique structure by some combination of intrinsic physicochemical constraints and evolutionary selection pressure. Simple exact models of proteins have suggested that accumulation of mutations could in principle cause a protein to switch to another fold, but also that optimization for folding in one form tends to steer evolution away from transitional sequences that could lead to a new structure.[1]

Nonetheless, some findings support a higher level of structural plasticity in proteins and less robust structural specificity.[2–5] Some natural proteins switch fold as a function of solution conditions, oligomerization or posttranslational modification.[6–10] Design and selection experiments have encoded two protein folds in a single amino-acid sequence or in nearly identical sequences.[11–14] Yadid *et al*. recently evolved fragments of tachylectin-2, a monomer, into pentamers that contain three different backbone topologies.[14] Limited mutagenesis of small natural proteins populates alternate folds in certain cases,[15–17] and accumulation of simple mutations can lead to

Abbreviations: CID, collision-induced dissociation; nESI-MS, nano-electrospray ionization mass spectrometry.

Additional Supporting Information may be found in the online version of this article.

Eric D. Dodds's current address is Department of Chemistry, University of Nebraska, Lincoln, NE 68588-0304, USA.
Vicki H. Wysocki's current address is Department of Chemistry and Biochemistry, Ohio State University Columbus, OH 43210-1173, USA.

Grant sponsor: National Institute of Health; Grant numbers: GM066806 (R01 to M.H.J.C.) and GM008804 (T32 to K.L.S.); Grant sponsor: NSF; Grant number: DBI 0923551 (to V.H.W.); Grant sponsor: University of Nebraska and Nebraska Tobacco Settlement Biomedical Research Development Fund (to E.D.D.).

*Correspondence to: Matthew H. J. Cordes, Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ. E-mail: cordes@email.arizona.edu

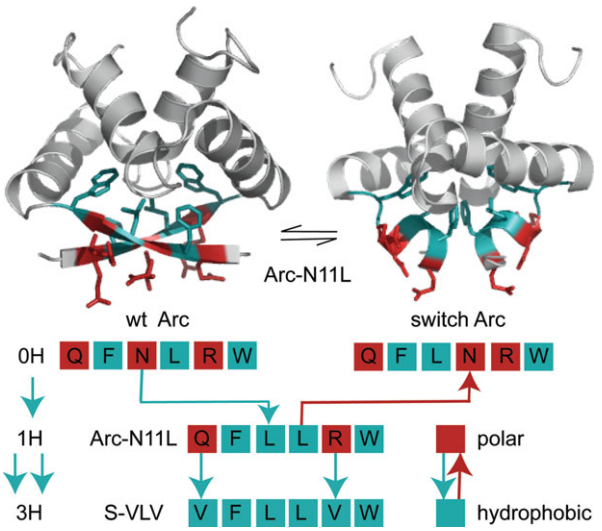

Figure 1. Hydrophobicity mutations and structural changes in the β-sheet of Arc repressor. In the wild-type Arc repressor homodimer (top left), residues 9–14 of each subunit alternate between polar (red) and hydrophobic (cyan) residues and interact to form a two-stranded β-sheet. A hydrophobic mutation at residue 11 (Arc-N11L) populates an alternate fold in which the β-sheet is replaced by 3_{10} helices (top right; determined in the switch Arc variant, which forms this fold exclusively). Two additional polar-to-hydrophobic substitutions produce a fully hydrophobic strand region in Arc-VLV.

evolution of new folds.[17–19] Some modeling studies have found numerous close approaches or paths between different structures in sequence space[20–23] and even sequence "supernetworks"[24,25] that span many protein folds. In addition, sequences with two folds or functions may confer a fitness advantage under certain models of adaptive evolution.[26] Yet the hypotheses of extensive connectedness and common evolutionary flow between sets of sequences compatible with different folds remain unconfirmed.

The sequence determinants of structural specificity and switching are underexplored,[27] but hydrophobicity is probably one major factor.[28] Protein folding is driven in large part by hydrophobic association, and protein sequences generally require a threshold level of hydrophobicity to fold stably into a compact structure,[29–32] but simple protein models suggest that excessively hydrophobic sequences are less likely to fold uniquely.[28,33–35] High or increased hydrophobicity is experimentally associated with chameleon sequences, lowered structural specificity, or outright fold changes in a number of proteins.[15,18,36–38] Effects can range from increased dynamics and loss of tertiary specificity without a clear change in fold,[36,37] to native state aggregation,[39] to variations in tertiary folding topology,[15] to nonnative aggregation including amyloidogenesis.[40–42]

In previous work (Fig. 1), a single polar-to-hydrophobic substitution (N11L) in the Arc repressor homodimer led to population of an alternate dimeric fold in which two 3_{10} helices replaced a two-stranded β-sheet.[15] A second Arc variant known as switch Arc (N11L/L12N) formed the alternate dimer structure exclusively, allowing determination of its structure.[43,44] While residue 11 is on the solvent-exposed face of the β-sheet in the native dimer, it flips into the protein's interior in the alternate dimer. Addition of surface hydrophobic residues can thus lower structural specificity by stabilizing alternate forms that bury the new hydrophobic surface area. This study showed that a single substitution could populate an alternate native structure, and confirmed theoretical predictions that surface polar-to-hydrophobic mutations can increase fold degeneracy.[45]

The same series of studies included variants in which all three surface polar residues in the β-sheet were replaced by hydrophobic residues.[43] One such variant, Q9V/N11L/R13V (S-VLV), included the N11L mutation plus two valine substitutions (Fig. 1). S-VLV showed an unusual thermal denaturation profile and appeared to aggregate, but was not studied in detail. We will now show that, while the single hydrophobic substitution in N11L populates an alternative dimeric fold with increased helical content, the two additional substitutions in S-VLV preserve both dimeric folds but also populate a thermally stable octamer with about half the helicity of the wild-type dimer. Thus, a single sequence can encode at least three folds with wide-ranging secondary structure content, and progressive increases in local sequence hydrophobicity progressively reduce fold specificity. We also suggest that switches in native folding topology involving oligomerization can be part of a near continuum of structural changes that includes amyloid aggregation.

Results and Discussion

S-VLV has two dimeric folds

We overexpressed S-VLV in histidine-tagged form, purified the protein by denaturing nickel affinity chromatography, and refolded the protein by dialysis into 50 m*M* MES (pH 5.5), 50 m*M* KCl. Size exclusion chromatograms of S-VLV at 300 μ*M* loading concentrations (Fig. 2a) showed both an apparent dimer (~15 kDa) and a second oligomer, larger than dimer but smaller than the column exclusion limit of ~100 kDa, indicated by the void volume V_0. By contrast, we observed only dimer for comparable preparations of wild-type Arc, Arc-N11L, and switch Arc. MALDI spectra of the higher and lower molecular weight size exclusion fractions of S-VLV give identical subunit masses (7600.9 Da), confirming that they represent different oligomeric states of the same protein. In addition, the two separated forms undergo slow exchange over hours to days as measured by size exclusion. ^{15}N-^{1}H NMR correlation spectra (Fig. 2b) of size-exclusion purified, uniform ^{15}N-labeled VLV dimer are dominated by signals from the dimer,

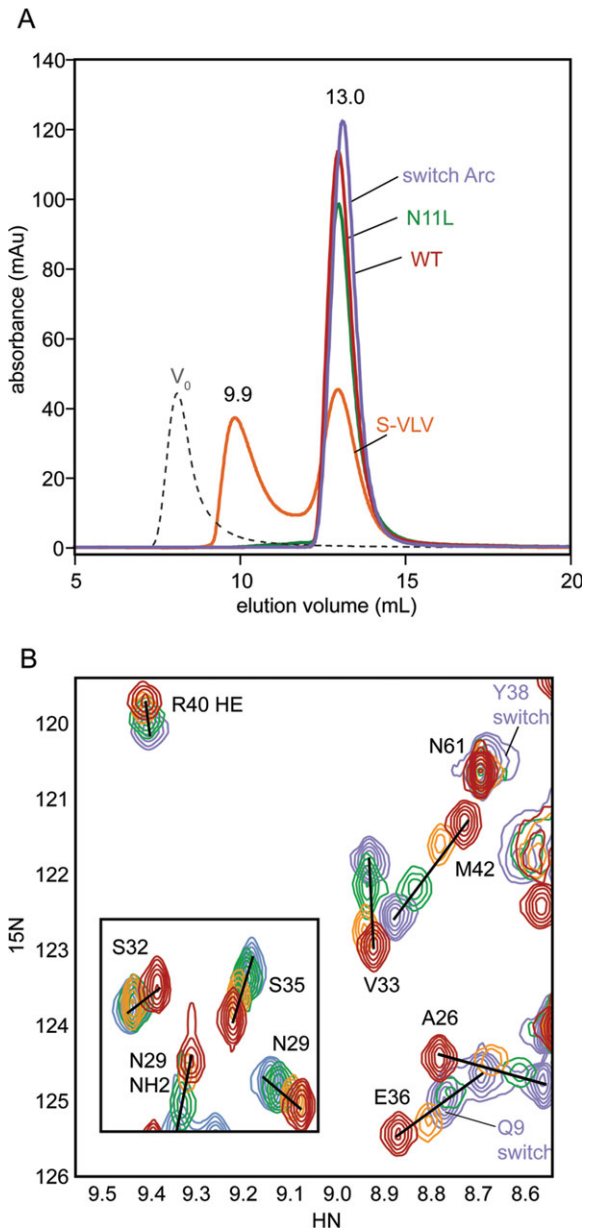

Figure 2. (A) Size exclusion chromatograms of Arc-S-VLV (orange), switch Arc (lavender), Arc-N11L (green), and Arc-WT (red), at 300 μ*M* protein concentration in 50 m*M* MES (pH 5.5), 50 m*M* KCl. Switch Arc, Arc-N11L, and WT migrate as dimers with elution volumes of 13.0 ± 0.1 mL, while Arc-S-VLV migrates as a mixture of dimer (13.0 mL) and higher-order oligomer (9.9 mL). V_0 (dashed line) indicates the void volume of the column, measured using blue dextran. B) Overlay of HSQC spectra for the four Arc variants at 25°C, colored as in panel A. Black lines track chemical shift changes for equivalent resonances. For all resonances shown, Arc-N11L and S-VLV have intermediate chemical shift values, indicative of two different Arc dimer folds in fast conformational exchange.

presumably due to rapid transverse relaxation and slow repopulation of the higher order oligomer.

The S-VLV dimer spectra show characteristics similar to those of Arc-N11L, indicating the presence of both dimeric Arc folds. Previous NMR lineshape studies on Arc-N11L[15] showed relatively rapid exchange at 25°C (k_{ex}=2200 ± 500 s^{-1} under slightly different solution conditions) between the β-sheet and 3_{10}-helical dimeric folds, both of which are significantly populated. For residues with small but significant chemical shift differences between wild-type and switch Arc, the resonances in Arc-N11L are in fast exchange between the two folds, and each shows a single peak with an intermediate chemical shift value (e.g. Asn 29, Ser 32 and Ser 35, inset in Fig. 2b). Resonances with larger chemical shift changes are on the edge of the intermediate exchange regime and show limited exchange broadening with attenuation of signal intensity (e.g. Ala 26 and Val 33). Significantly, equivalent resonances of the S-VLV dimer show very similar intermediate behavior except that the resonances are closer to the positions observed in wild-type Arc, while the resonances in Arc-N11L are closer to the switch Arc positions (Fig. 2b). We conclude that S-VLV, like Arc-N11L, exists in two dimeric folds that exchange on the millisecond to microsecond time scale; in addition, the resonance positions suggest that the equilibrium in S-VLV favors the β-sheet fold by about 3:1, while the equilibrium in Arc-N11L favors the helical fold by a similar ratio.

S-VLV also forms an octamer

When refolded in a different, higher ionic strength buffer [SB250 buffer; 50 m*M* Tris (pH 7.5), 250 m*M* KCl, 0.2 m*M* EDTA], S-VLV almost exclusively formed a higher order oligomer, and any residual dimer could be removed by a heat annealing treatment at 80°C. After the heat annealing treatment the oligomer remained stable over time. By analytical size exclusion chromatography (Fig. 3), the heat-annealed S-VLV gave an elution volume consistent with a molecular weight of 65 kDa, much larger than the wild-type Arc homodimer (15.0 kDa, apparent molecular weight) and suggestive of an octamer or nonamer.

Sedimentation equilibrium analysis of heat-treated S-VLV (Supporting Information Fig. S1) suggests an octamer. We fitted radial absorbance curves to models that assume a single ideal species, across a three fold range of loading concentrations and three rotor speeds. At 30 μ*M* S-VLV protein, average apparent molecular weight ($M_{w,app}$) for three rotor speeds was 59 kDa, or 7.8 subunits; values for 60 μ*M* and 90 μ*M* protein were 57 kDa (7.5 subunits) and 57 kDa, (7.5 subunits), respectively. The sedimentation data are consistent with a single octameric species but do not exclude a mixture of multiple species of similar oligomeric state, for example, a heptamer and octamer. Nor can we rigorously exclude such a mixture on the basis of the observation of a single peak in the size exclusion trace.

Native nano-electrospray ionization mass spectrometry (nESI-MS) of heat-treated S-VLV in 0.1*M* ammonium acetate (pH 7.4) confirmed the presence

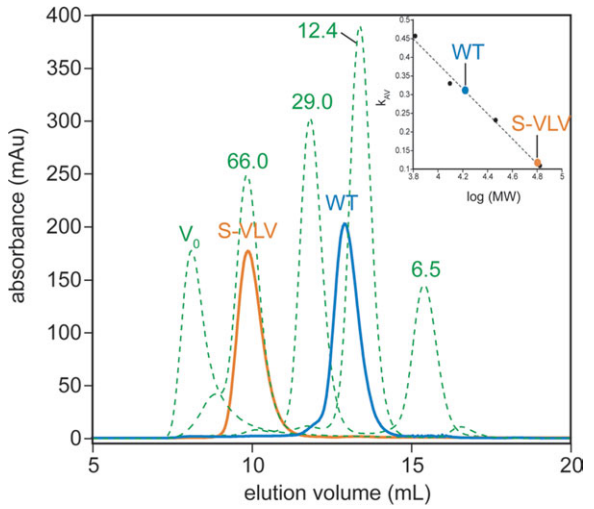

Figure 3. Size exclusion chromatograms of wild-type Arc (blue solid trace) and heated S-VLV (orange solid trace), both at 350 μ*M* loading concentration, along with calibration standards (green dashed trace) labeled with molecular weights in kDa. The column void volume (V_0) is measured as the elution volume of blue dextran. The inset shows a calibration plot for estimation of molecular weight of S-VLV (65 kDa based on the regression) and wild-type Arc (15 kDa based on the regression, with an actual dimeric size of 15.4 kDa).

of a specifically octameric species (Fig. 4). Peaks observed at mass-to-charge ratios (m/z) > 3,500 corresponded to the charge state envelope of a distinct octamer (charge deconvoluted M_w of 60.8216 ± 0.0028 kDa, or 8.00309 ± 0.00036 subunits; mean ± 99% confidence interval for $n = 5$ measurements) rather than a mixture of multiple higher oligomeric states. The narrow charge state envelope of the octamer (only three charge states observed) and the octamer charge states themselves (which fall below the Rayleigh-limited maximum charge of $z = 20$ predicted for a native complex of this molecular weight) are consistent with an ordered, native-like structure.[46] Tandem mass spectrometry (MS/MS) via collision-induced dissociation (CID) further corroborated the octameric stoichiometry of the high m/z signals, with the octamer dissociating to yield monomer (charge deconvoluted M_w of 7.60068 kDa, or 1.00012 subunits) and the complementary heptamer (charge deconvoluted M_w of 53.2226 kDa, or 7.00319 subunits) in accord with the typical CID behavior of protein complexes.[47] The high m/z peaks observed for S-VLV were completely absent in the mass spectrum of wild-type Arc. Both wild-type Arc and S-VLV spectra also showed peaks originating from dimeric and monomeric forms. Size exclusion traces of S-VLV in the ammonium acetate buffer indicated a small population of dimeric S-VLV under these conditions.

Surprisingly, nESI-MS spectra of S-VLV in the absence of heat treatment show mostly heptamer rather than octamer (Supporting Information Figs. S2 and S3). Sedimentation equilibrium runs of unheated samples also yield reduced apparent molecular weights largely consistent with heptamer (Table SI), though greater populations of dimer in such samples could also produce this effect. Size exclusion traces of unheated and heated samples have no systematic variation in elution volume within experimental error. However, the clear appearance of heptamer and octamer in nESI-MS of unheated and heat-treated samples, respectively, indicates that S-VLV can fold into multiple oligomers of similar size and suggests that an octamer is the most thermodynamically stable quaternary structure. Slow formation of a very stable octamer upon heating may also explain why heat-treated samples of S-VLV are less prone to repopulation of dimer.

The octamer is highly stable and differently folded from either dimer

We next analyzed the secondary structure content of heat-treated S-VLV oligomers using far ultraviolet circular dichroism (Fig. 5a). Wild-type Arc has two α-helices per monomer and approximately 50% total

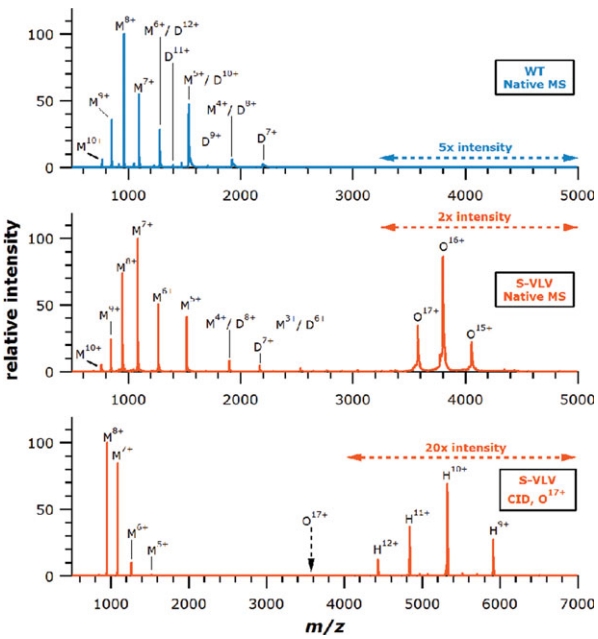

Figure 4. Native nESI-MS of wild-type Arc and S-VLV, each at 50 μ*M* protein in 0.1*M* ammonium acetate (pH 7.4), and CID of the S-VLV octamer. Under native MS conditions, wild-type Arc shows peaks for monomeric (M) and dimeric (D) states, while S-VLV contains peaks for monomer, dimer, and octamer (O); note that relative intensities do not directly reflect relative abundances of different oligomeric states because of instrument discrimination at higher m/z. CID of the S-VLV octamer (z = 17) produced highly charged monomer (consistent with the commonly expected oligomer CID pathway that involves unfolding and release of one subunit) and complementary heptamer (H). Charge states are indicated by superscripts.

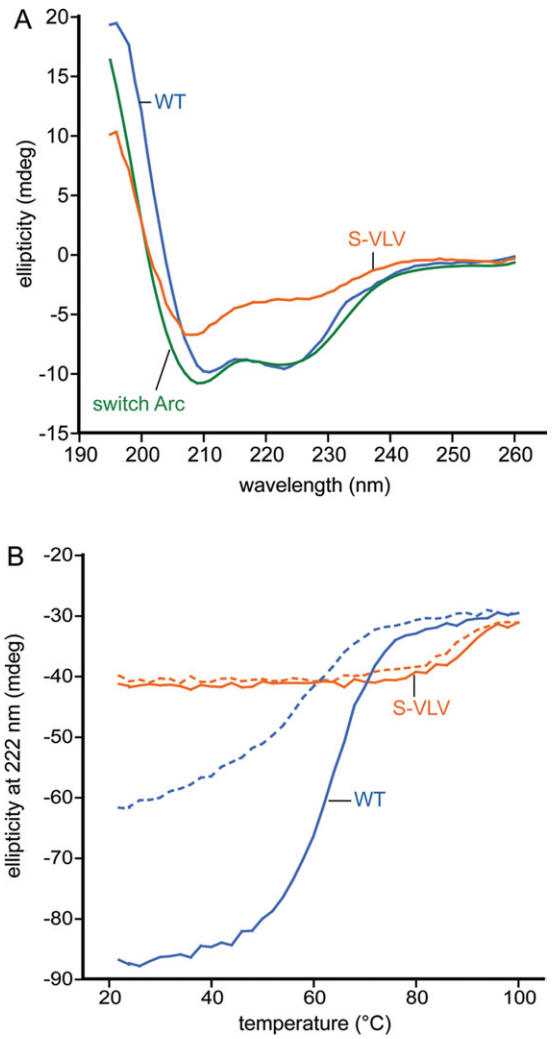

Figure 5. (A) Far UV circular dichroism spectra of wild-type Arc (blue), heated S-VLV (orange), and switch Arc (green) in SB250 (100 μ*M* protein, 0.1 mm pathlength, 20°C), B) Thermal denaturation of wild-type Arc (blue) and heated S-VLV (orange) monitored at 222 nm (50 μ*M* protein, 2.0 mm pathlength), with forward melts shown as solid lines and reverse melts shown as dashed lines.

helical content based on the native sequence plus the affinity tag.[48] Signals characteristic of helical structure (208 nm and 222 nm) were less intense for S-VLV than for wild-type Arc, with the 222 nm signal in particular being about half as strong, suggesting major loss of α-helical content. The spectrum of S-VLV does not, however, have the strong signal at 200 nm expected for a disordered or unfolded structure. Secondary structure analysis using CONTINLL[49] and SELCON3[50] (Table SII) gave an average of 56% helical content for wild-type Arc and 32% helical content for S-VLV, consistent with loss of one of the two helices present in the wild-type structure or parts of both. These methods also calculate that the lost helical content is replaced by a mixture of absolute increases in β-strand (12%), turn (5%) and disordered content (5%), indicating that S-VLV is folded, but very differently both from the wild-type Arc dimer and the alternate dimer switch Arc (Figs. 1 and 5A). Near ultraviolet spectra (Supporting Information Fig. S4) confirm significant but distinct tertiary and quaternary interactions for S-VLV.

To investigate the folding stability of the S-VLV octamer, we monitored thermal denaturation by circular dichroism at 222 nm (Fig. 5b). The dynamic range of the S-VLV melt was reduced relative to wild-type Arc, due to the lower magnitude ellipticity of S-VLV at 222 nm. Surprisingly, the thermal denaturation midpoint (T_m) of S-VLV occurred near 90°C, more than 25°C higher than that of wild-type Arc at the same protein concentration (50 μ*M*). S-VLV unfolded and refolded reversibly, while wild-type Arc refolded only partially in a reverse melt under these conditions. Thus, the structure of S-VLV exhibits remarkable ability to refold and resistance to thermal denaturation.

Overlap of multiple stable folds in sequence space

We have shown that accumulation of polar-to-hydrophobic substitutions in the Arc repressor β-sheet first populates a more helical dimer with one mutation (Arc-N11L), and then a much less helical higher-order oligomer with three (S-VLV).[15,44] In S-VLV, all three forms coexist under the same solution conditions (Fig. 2). The oligomer is not a nonspecific aggregate: it is an octamer following a heat annealing treatment, though a heptamer may initially form under kinetic control during refolding by dialysis. The heat-annealed S-VLV octamer is at least partially ordered and highly stable and appears to form a cooperatively folded unit. The low helicity strongly suggests that some if not all of the individual subunits fold differently from the predominantly helical subunits in either dimer; thus, the octamer does not derive trivially from simple self-association of dimers but also contains changes in secondary structure content. Thus, S-VLV has at least three folds that represent an extreme level of protein structural plasticity and loss of fold specificity with respect to simple changes in amino-acid sequence. The detailed structures of the octamer and heptamer are unknown at present but are the subject of ongoing NMR and crystallographic studies. Their slow exchange with each other and with dimer suggests large topological differences that can interconvert only through global unfolding. We speculate that the higher oligomers may contain a β–barrel or sandwich core, which serves to bury the newly introduced hydrophobic side chains at positions 9, 11, and 13 that would otherwise remain solvent-exposed in the dimer.

This work shows that multiple (more than two) stable folds may be connected by overlapping regions in sequence space, and promotes the view that evolution of new folds from preexisting folds would not

be strongly hindered by a requirement for continuous preservation of folding stability.[5,18,51] Sequences that can encode two stable folds have been referred to as 'bridges', 'switches', 'metamorphic proteins' and 'transformer proteins' among other terms.[2,15,52–54] S-VLV is in effect a three-way bridge or three-fold switch,[53] capable in principle of acting as a hub for mutational traffic between at least three stable folding patterns related by nontrivial changes in secondary structure. In a related but distinct example, Yadid *et al.*[14] recently described a sequence with three subunit folding topologies within a single folded oligomeric structure.[14] Proteins that exhibit radical structural polymorphism, involving simultaneous or successive population of more than two specific folded backbone topologies in a single sequence or a series of variants, might reasonably be called 'polymetamorphic proteins'. Dill[35] referred to sequences with multiple native structures as 'gemisch' sequences, but this usage was clearly intended to encompass the detailed specificity of the native structure as well as large shifts in folding topology; thus, a protein with an intrinsically disordered or 'molten' native state might also be a gemisch sequence.

The three hydrophobic mutations appear to increase the overall folding stability of Arc even as they lower the fold specificity. Arc-N11L, a two-way bridge sequence, has a higher melting temperature than wild-type Arc;[17] S-VLV, a three-way bridge sequence, has a higher melting temperature than Arc-N11L. The multimerization of S-VLV may provide a mechanism for the protein to maintain and even increase stability while undergoing structural rearrangement.[55] In general, however, it may be more common for bridge sequences to have lower native state stability than sequences with a single native structure.[23]

Sequence hydrophobicity and folding degeneracy

The sequence determinants of native state conformational specificity are not well understood, despite an eloquent call by Lattman and Rose for efforts in this direction nearly two decades ago.[27] Some previous computational and experimental work points to the importance of hydrophobicity. Simple exact models predict that proteins with a unique, compact native state will have intermediate sequence hydrophobicity, because low hydrophobicity hinders compactness and high hydrophobicity increases structural degeneracy.[33,45] Hydrophobic mutations to proteins do reduce structural specificity in some cases, producing a molten, gemisch state[36,37] or even an alternate fold in the case of Arc-N11L.[15] The behavior of S-VLV now shows for the first time that progressive increases in hydrophobicity lead to progressively higher structural degeneracy in the form of multiple different folding patterns. Although substitution of either buried[37] or surface[36] polar residues can compromise conformational specificity, the results presented here emphasize surface mutations.

Lattman and Rose suggested that the determinants of specificity could either be distributed across the entire sequence or centralized in certain "tender spots", alteration of which would yield a conformational "catastrophe".[27] The Arc repressor variants are more consistent with centralized control of specificity, in that the progressive population of alternate topologies results from multiple mutations within a short stretch of sequence. Concentrated hydrophobic substitutions in other regions of Arc repressor do not clearly alter the fold;[56] the effect of dispersed sets of mutations in this system remains to be assessed.

High local sequence hydrophobicity also promotes amyloid aggregation.[40] In S-VLV the localized hydrophobic mutations lead to a continuous hydrophobic stretch from 9 to 14 (VFLLVW). TANGO[57] predicts this region to be highly amyloidogenic in S-VLV but not in N11L or wild-type Arc. In the experiments reported here we found no evidence of highly aggregated forms of S-VLV, neither precipitates nor soluble aggregates that run at the void volume of the size exclusion column. Extensive hydrophobic aggregation in S-VLV may be prevented by proline residues that flank the putative amyloidogenic region at positions 8 and 15. Proline residues are known to act as "gatekeepers" or amyloid breakers.[58]

Is the behavior of S-VLV an arrested form of amyloid aggregation? Amyloid fibril formation can proceed through a series of oligomeric precursors; under some conditions, these species have been isolated, characterized and found to have specific structures, such as the recently identified "cylindrin" hexamers for αβ crystallin.[59] Both cylindrins and S-VLV assemble by topological rearrangement of the starting structure into stable, discrete oligomers with β character. Cylindrins and other amyloid precursor oligomers are generally cytotoxic and can be recognized by precursor-specific antibodies.[60] Neither of these properties has yet been tested for S-VLV, and it remains possible that the S-VLV oligomers are akin to a particularly stable, long-lived amyloid precursor species.

The above comparisons hint at a unified view in which reduced protein fold specificity manifests as a near continuum of topological possibilities. These range from defined alterations in native folding pattern with varying degrees of change, to rearrangements coupled with specific or nonspecific oligomerization, to open-ended nonnative self-association including amyloid aggregation. These different structural changes may share some sequence determinants, such as stretches of hydrophobic residues; which structural changes (if any) are observed for a particular sequence variation may depend both on the context and on the degree of sequence perturbation. Very hydrophobic stretches of sequence are disfavored in protein evolution.[61,62] This avoidance has been

attributed to selection against amyloid aggregation, but may also relate to a general avoidance of reduced fold specificity, of which amyloidogenesis is an extreme and especially hazardous outcome.

Materials and Methods

Mutagenesis, protein expression and affinity purification

A plasmid encoding the Arc repressor triple mutant S-VLV was constructed by QuikChange (Stratagene; Santa Clara, CA) site-directed mutagenesis of the synthetic *arc-st11* gene in plasmid pET800.[63] The *arc-st11* gene includes the coding sequence for the st11 C-terminal extension (H_6KNQHE) to stabilize the proteins against degradation during expression and to allow for affinity purification.[64] Wild-type Arc and S-VLV proteins were overexpressed in *Escherichia coli* strain BL21(λDE3) cells transformed with the appropriate variant of pET800, and purified by Ni^{2+}-NTA affinity chromatography under denaturing conditions by loading cleared cell lysates in 100 m*M* NaH_2PO_4 (pH 8.0), 10 m*M* Tris, 6*M* guanidine hydrochloride, 10 m*M* imidazole onto 3 mL Ni^{2+}-NTA resin per liter of culture, washing with 40, 40, and 20 mL fractions of the same lysis/loading buffer, and eluting with low pH using 6*M* guanidine hydrochloride containing 0.2*M* acetic acid. After elution, proteins were refolded by dialysis into SB250 buffer [50 m*M* Tris (pH 7.5), 250 m*M* KCl, 0.2 m*M* EDTA] for 48 h at 4°C, with one buffer change after 24 h.

NMR spectroscopy

For structural comparisons of S-VLV, Arc-N11L, wild-type and switch Arc, expression plasmids were transformed into *Escherichia coli* strain BL21(λDE3) cells and grown in M9T minimal media with 0.8 g/L ^{15}N-labelled $^{15}NH_4Cl$ as sole nitrogen source. Proteins were purified by Ni^{2+}-NTA affinity chromatography under denaturing conditions as described above, refolded by dialysis into 50 m*M* MES (pH 5.5), 50 m*M* KCl, and further purified by size exclusion chromatography in the same buffer with a Superdex 75 HR 10/30 column (GE Healthcare). Wild-type Arc, Arc-N11L, and switch Arc each eluted as single peaks at 13.0 mL, while S-VLV eluted in two peaks corresponding to dimer (13.0 mL) and a higher oligomer (9.9 mL). Dimer-containing fractions for all four proteins were diluted to a low protein concentration (87 μ*M*) to allow maintenance of the dimeric state in S-VLV, and brought to 10% D_2O (v/v) and 0.01% sodium azide (w/v). NMR experiments were performed on a Varian Inova 600 MHz spectrometer equipped with a $^{1}H/^{15}N/^{13}C$ cryogenic probe. ^{15}N-^{1}H correlation (HSQC) spectra were collected in 5°C increments from 5 to 25°C. Data were processed with NMRPipe (NIH) and analyzed with SPARKY (UCSF).

Size exclusion chromatography

Affinity-purified, refolded wild-type Arc and S-VLV proteins were injected onto a Superdex 75 size exclusion column (HR 10/30, GE Healthcare) at 350 μ*M* loading concentrations and eluted isocratically with SB250 buffer. Two sharp peaks were typically visible in the resulting S-VLV chromatogram: a large peak at 9.9 mL elution volume, representing oligomeric S-VLV, and a smaller peak (up to approximately 10% the integrated area of the larger peak) at 13.0 mL, representing dimeric S-VLV. Heating the S-VLV samples to 80°C for 2 h and cooling prior to injection eliminated the apparent dimer, possibly due to conversion of protein initially refolded as dimer to a more thermodynamically stable octamer. When heat-annealed S-VLV was loaded onto the size exclusion column, only the putative oligomer peak was visible. The size exclusion fraction (1 mL) containing this peak was collected and centrifugally concentrated (Amicon Ultra) for use in further analyses. All subsequent experiments, unless otherwise noted, were performed with heat annealed, size-exclusion purified S-VLV. The void volume of the Superdex column was determined using the elution volume of blue dextran (1 mg/mL; 2000 kDa) (Sigma-Aldrich). The molecular weight of S-VLV was estimated by a linear regression analysis using the elution volumes of low molecular weight protein calibration standards (Sigma-Aldrich) including aprotinin (3 mg/mL; 6.5 kDa), cytochrome c (2 mg/mL; 12.4 kDa), carbonic anhydrase (2 mg/mL; 29.0 kDa), bovine serum albumin (5mg/mL; 66.0 kDa). All standards and protein samples were injected at 0.5 mL volume.

Sedimentation equilibrium

S-VLV protein at three loading concentrations (30 μ*M*, 60 μ*M*, and 90 μ*M*) in SB250 was analyzed at three rotor speeds (10000 rpm, 13000 rpm, and 15000 rpm) at 20°C in a Beckman XL-I analytical ultracentrifuge to generate radial distribution curves. Absorbance was monitored at 280 nm with a radial spacing of 0.001 cm. At each rotor speed, 20 replicate scans were collected. The apparent molecular weight of S-VLV was determined by fitting sedimentation curves to a standard single species model using Kaleidagraph (Synergy Software; Reading, PA). Buffer density and partial specific volume were estimated using SEDNTERP (J. Philo; Thousand Oaks, CA) The apparent molecular weight estimated in Kaleidagraph was used to determine the apparent number of subunits using the isotopically averaged molecular weight for st11-tagged S-VLV (7599.7 Da).

Mass spectrometry

Wild-type Arc and S-VLV proteins were refolded and dialyzed into 1.0*M* ammonium acetate (pH 7.0) following Ni^{2+}-NTA affinity purification. Each sample was

diluted to 50 μ*M* concentration and buffer exchanged into 0.1*M* ammonium acetate, pH 7.4, using size exclusion spin columns (BioRad, Hercules, CA). Mass spectra were acquired using a Waters Synapt G2 quadrupole time-of-flight instrument (Manchester, UK). Typical operating conditions for analysis of protein complexes using this instrument have been described in detail recently.[65] Briefly, native nESI was performed using home-pulled borosilicate emitters loaded with 10 μL of buffer exchanged protein solution and fitted onto a custom nESI source. Instrument parameters were optimized for preservation of non-covalent protein complexes. Most critically, the inlet pressure was raised to approximately five times the normal level to provide collisional ion cooling, while DC offsets throughout the instrument were reduced to the lowest practical values to prevent unwanted ion activation. Charge state envelopes were deconvoluted to neutral masses using MagTran.[66]

Circular dichroism

Far ultraviolet circular dichroism spectra of wild-type Arc, S-VLV, and switch Arc were obtained in SB250 at 20°C using 100 μ*M* size exclusion-pure protein in a 0.1 mm path length cuvette on a Olis DSM-20 CD spectrometer. Scans were collected using a 10 second integration time per scan and signal averaging of three scans measured at 1 nm increments from 260 to 195 nm. Thermal denaturation profiles of wild-type Arc and S-VLV were obtained in SB250 using 50 μ*M* size exclusion-pure protein samples in a 2.0 mm pathlength cuvette. Samples were heated from 20 to 100°C with a 3 minute sample equilibration time, 2°C step size, and 60 s data collection time while ellipticity was monitored at 222 nm. Reverse melts utilized the same parameters as the forward melts. Near UV circular dichroism spectra of wild-type Arc, S-VLV, and switch Arc were obtained in SB250 at 20°C using 50 μ*M* size exclusion-purified protein in a 1.0 cm pathlength cuvette. Scans were collected using a 20 second integration time per scan and signal averaging of three scans measured at 1 nm increments from 305 to 260 nm.

REFERENCES

1. Chan HS, Bornberg-Bauer E (2002) Perspectives on protein evolution from simple exact models. Appl Bioinformatics 1:121–144.
2. Murzin AG (2008) Metamorphic proteins. Science 320: 1725–1726.
3. Bryan PN, Orban J (2010) Proteins that switch folds. Curr Opin Struct Biol 20:482–488.
4. Cordes MHJ, Stewart KL (2012) The porous borders of the protein world. Structure 20:199–200.
5. Meier S, Ozbek S (2007) A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution? Bioessays 29:1095–1104.
6. Knauer SH, Rosch P, Artsimovitch I (2012) Transformation: the next level of regulation. RNA Biol 9:1418–1423.
7. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I, Rosch P (2012) An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 150:291–303.
8. Luo XL, Tang ZY, Xia GH, Wassmann K, Matsumoto T, Rizo J, Yu HT (2004) The Mad2 spindle checkpoint protein has two distinct natively folded states. Nat Struct Mol Biol 11:338–345.
9. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci USA 105:5057–5062.
10. Littler DR, Harrop SJ, Fairlie WD, Brown LJ, Pankhurst GJ, Pankhurst S, DeMaere MZ, Campbell TJ, Bauskin AR, Tonini R, Mazzanti M, Breit SN, Curmi PMG (2004) The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. J Biol Chem 279:9298–9305.
11. Ambroggio XI, Kuhlman B (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds. J Am Chem Soc 128:1154–1161.
12. He YA, Chen YH, Alexander PA, Bryan PN, Orban J (2012) Mutational tipping points for switching protein folds and functions. Structure 20:283–291.
13. Ambroggio XI, Kuhlman B (2006) Design of protein conformational switches. Curr Opin Struct Biol 16: 525–530.
14. Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS (2010) Metamorphic proteins mediate evolutionary transitions of structure. Proc Natl Acad Sci USA 107: 7287–7292.
15. Cordes MHJ, Burton RE, Walsh NP, McKnight CJ, Sauer RT (2000) An evolutionary bridge to a new protein fold. Nat Struct Biol 7:1129–1132.
16. Tidow H, Lauber T, Vitzithum K, Sommerhoff CP, Rosch P, Marx UC (2004) The solution structure of a chimeric LEKTI domain reveals a chameleon sequence. Biochemistry 43:11238–11247.
17. Meier S, Jensen PR, David CN, Chapman J, Holstein TW, Grzesiek S, Ozbek S (2007) Continuous molecular evolution of protein-domain structures by single amino acid changes. Curr Biol 17:173–178.
18. Anderson WJ, Van Dorn LO, Ingram WM, Cordes MHJ (2011) Evolutionary bridges to new protein folds: design of C-terminal Cro protein chameleon sequences. Protein Eng Des Sel 24:765–771.
19. Roessler CG, Hall BM, Anderson WJ, Ingram WM, Roberts SA, Montfort WR, Cordes MH (2008) Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. Proc Natl Acad Sci USA 105:2343–2348.
20. Babajide A, Farber R, Hofacker IL, Inman J, Lapedes AS, Stadler PF (2001) Exploring protein sequence space using knowledge-based potentials. J Theor Biol 212:35–46.
21. Meyerguz L, Kleinberg J, Elber R (2007) The network of sequence flow between protein structures. Proc Natl Acad Sci USA 104:11627–11632.
22. Cao BQ, Elber R (2010) Computational exploration of the network of sequence flow between protein structures. Proteins 78:985–1003.
23. Sikosek T, Bornberg-Bauer E, Chan HS (2012) Evolutionary dynamics on protein bi-stability landscapes can potentially resolve adaptive conflicts. PLoS Comput Biol 8:e1002659.
24. Burke S, Elber R (2012) Super folds, networks, and barriers. Proteins 80:463–470.
25. Cui Y, Wong WH, Bornberg-Bauer E, Chan HS (2002) Recombinatoric exploration of novel folded structures:

a heteropolymer-based model of protein evolutionary landscapes. Proc Natl Acad Sci U S A 99:809–814.
26. Sikosek T, Chan HS, Bornberg-Bauer E (2012) Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. Proc Natl Acad Sci U S A 109:14888–14893.
27. Lattman EE, Rose GD (1993) Protein folding – What's the question? Proc Natl Acad Sci USA 90:439–441.
28. Sandelin E (2004) On hydrophobicity and conformational specificity in proteins. Biophys J 86:23–30.
29. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155.
30. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Prot Chem 14:1–63.
31. Pace CN, Fu HL, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Iimura S, Gajiwala K, Scholtz JM, Grimsley GR (2011) Contribution of hydrophobic interactions to protein stability. J Mol Biol 408: 514–528.
32. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41:415–427.
33. Chan HS, Dill KA (1991) Sequence space soup of proteins and copolymers. J Chem Phys 95:3775–3787.
34. Yue K, Dill KA (1992) Inverse protein folding problem – designing polymer sequences. Proc Natl Acad Sci USA 89:4163–4167.
35. Dill KA, Bromberg S, Yue KZ, Fiebig KM, Yee DP, Thomas PD, Chan HS (1995) Principles of protein-folding – A perspective from simple exact models. Protein Sci 4:561–602.
36. Hill RB, DeGrado WF (2000) A polar, solvent-exposed residue can be essential for native protein structure. Struct Fold Des 8:471–479.
37. Bolon DN, Mayo SL (2001) Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. Biochemistry 40:10047–10053.
38. Lumb KJ, Kim PS (1998) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. Biochemistry 37:13042.
39. Padlan EA, Love WE (1985) Refined crystal-structure of deoxyhemoglobin-S .2. Molecular-interactions in the crystal. J Biol Chem 260:8280–8291.
40. Dubay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. J Mol Biol 341:1317–1326.
41. Calamai M, Taddei N, Stefani M, Ramponi G, Chiti F (2003) Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. Biochemistry 42:15078–15083.
42. Kim W, Hecht MH (2006) Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's A beta 42 peptide. Proc Natl Acad Sci USA 103: 15824–15829.
43. Cordes MHJ, Walsh NP, McKnight CJ, Sauer RT (2003) Solution structure of switch Arc, a mutant with 3(10) helices replacing a wild-type beta-ribbon. J Mol Biol 326:899–909.
44. Cordes MHJ, Walsh NP, McKnight CJ, Sauer RT (1999) Evolution of a protein fold in vitro. Science 284: 325–327.
45. Lau KF, Dill KA (1990) Theory for protein mutability and biogenesis. Proc Natl Acad Sci USA 87:638–642.
46. Heck AJR, van den Heuvel RHH (2004) Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23:368–389.
47. Benesch JLP (2009) Collisional activation of protein complexes: picking up the pieces. J Am Soc Mass Spectrom 20:341–348.
48. Raumann BE, Rould MA, Pabo CO, Sauer RT (1994) DNA recognition by beta-sheets in the Arc repressor-operator crystal-structure. Nature 367:754–757.
49. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular-dichroism. Anal Biochem 209:32–44.
50. Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular-dichroism. Biochemistry 20:33–37.
51. Van Dorn LO, Newlove T, Chang SM, Ingram WM, Cordes MHJ (2006) Relationship between sequence determinants of stability for two natural homologous proteins with different folds. Biochemistry 45: 10542–10553.
52. Yeates TO (2007) Protein structure: evolutionary bridges to new folds. Curr Biol 17:R48–R50.
53. Bornberg-Bauer E (1997) How are model protein structures distributed in sequence space? Biophys J 73: 2393–2403.
54. Knauer SH, Artsimovitch I, Rosch P (2012) Transformer proteins. Cell Cycle 11:4289–4290.
55. Frank MK, Dyda F, Dobrodumov A, Gronenborn AM (2002) Core mutations switch monomeric protein GB1 into an intertwined tetramer. Nat Struct Biol 9:877–885.
56. Cordes MH, Sauer RT (1999) Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. Protein Sci 8:318–325.
57. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22:1302–1306.
58. Rousseau F, Serrano L, Schymkowitz JWH (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. J Mol Biol 355:1037–1047.
59. Laganowsky A, Liu C, Sawaya MR, Whitelegge JP, Park J, Zhao ML, Pensalfini A, Soriaga AB, Landau M, Teng PK, Cascio D, Glabe C, Eisenberg D (2012) Atomic view of a toxic amyloid small oligomer. Science 335:1228–1231.
60. Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, Glabe CG (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science 300:486–489.
61. Schwartz R, King J (2006) Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. Protein Sci 15:102–112.
62. Patki AU, Hausrath AC, Cordes MHJ (2006) High polar content of long buried blocks of sequence in protein domains suggests selection against amyloidogenic nonpolar sequences. J Mol Biol 362:800–809.
63. Milla ME, Brown BM, Sauer RT (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor. Nat Struct Biol 1:830–830.
64. Milla ME, Brown BM, Sauer RT (1993) P22-Arc repressor – Enhanced expression of unstable mutants by addition of polar C-terminal sequences. Protein Sci 2:2198–2205.
65. Dodds ED, Blackwell AE, Jones CM, Holso KL, O'Brien DJ, Cordes MHJ, Wysocki VH (2011) Determinants of gas-phase disassembly behavior in homodimeric protein complexes with related yet divergent structures. Anal Chem 83:3881–3889.
66. Zhang ZQ, Marshall AG (1998) A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. J Am Soc Mass Spectrom 9:225–233.