Amino Acids (2012) 43:2231–2247
DOI 10.1007/s00726-012-1418-4

REVIEW ARTICLE

On the segregation of protein ionic residues by charge type

Michael S. Parker · Ambikaipakan Balasubramaniam · Steven L. Parker

Received: 4 October 2012 / Accepted: 6 October 2012 / Published online: 19 October 2012

Abstract Based on ubiquitous presence of large ionic motifs and clusters in proteins involved in gene transcription and protein synthesis, we analyzed the distribution of ionizable sidechains in a broad selection of proteins with regulatory, metabolic, structural and adhesive functions, in agonist, antagonist, toxin and antimicrobial peptides, and in self-excising inteins and intron-derived proteins and sequence constructs. All tested groups, regardless of taxa or sequence size, show considerable segregation of ionizable sidechains into same type charge (homoionic) tracts. These segments in most cases exceed half of the sequence length and comprise more than two-thirds of all ionizable sidechains. This distribution of ionic residues apparently reflects a fundamental advantage of sorted electrostatic contacts in association of sequence elements within and between polypeptides, as well as in interaction with polynucleotides. While large ionic densities are encountered in highly interactive proteins, the average ionic density in most sets does not change appreciably with size of the homoionic segments, which supports the segregation as a modular feature favoring association.

Keywords Ionic segregation · Homobasic tract · Homoacidic tract · Homoionic zipper

Abbreviations

aa	Amino acid (residue)
ec1	First extracellular domain
GPCR	G-protein coupling receptor
ic3	Third intracellular domain
ic4	Fourth intracellular domain
NLM	US National Library of Medicine
u-p	UniProtein database

Introduction

Ionizable amino acid sidechains, and especially the basic, are known to be considerably grouped by type of charge within sequences of taxonomically ubiquitous polynucleotide-handling proteins and chaperone/heat-shock proteins. The RNA-partnering ribosomal proteins and the DNA-engaging protamine- and histone-type proteins are well identified with long segments containing same-charge ionic sidechains (see e.g. [34] for ribosomal proteins, [7] for protamines, [58] for histones). Basic zippers are recognized in a large number of proteins [39, 86]. These segments have large electrostatic and hydrophobic interaction potentials, and could be considered as homoionic zippers. Cation detection by voltage-gated channels also depends on homoionic tracts [11], as does the nuclear localization of several protein classes [23, 24, 91]. A large residue

Electronic supplementary material The online version of this article (doi:10.1007/s00726-012-1418-4) contains supplementary material, which is available to authorized users.

M. S. Parker
Department of Microbiology and Molecular Cell Sciences, University of Memphis, Memphis, TN 38152, USA

A. Balasubramaniam
Department of Surgery, University of Cincinnati College of Medicine, Cincinnati, OH 45263, USA

S. L. Parker (✉)
Department of Pharmacology, University of Tennessee Health Sciences Center, Memphis, TN 38163, USA
e-mail: stevenleonardparker@msn.com

enrichment by type is also found in transmembrane (aliphatic and aromatic hydrophobic types) and intracellular (homoionic type) domains of membrane-resident proteins. The G-protein coupling receptors (GPCRs) are an example class in this regard (see [42, 62, 79]; and Table 6 in this survey).

Segregation of protein or peptide ionizable sidechains into homoionic tracts is not sufficiently explored. Polynucleotide-associating proteins (protamines and histones, transcription factors, ribosomal proteins and other nucleic acid-binding proteins) are implicitly accepted as possessing large local and global densities of basic residues. Calcium-binding motifs (e.g. in calmodulins and troponins) show large accumulations of acidic sidechains. Many types of peptidic agonists have segments rich in basic residues, and the cognate receptors (e.g. apelin [96], chemokine [73] and neuropeptide Y (NPY) [63]) have counter-matching anion-rich domains. Grouping of ionic residues could also be linked to structuring [20], favoring more open structure for the involved motifs [21, 47, 76, 89].

The above very diverse examples offer no compelling reason to assume that charge segregation within sequences would be absent from proteins at large. A random charge dispersion could be counterproductive for protein interaction in many paradigms, and especially in the primordial association with polynucleotides. Homoionic tracts obviously offer advantages in selection and engagement of a large variety of partners.

Homoionic tracts are indeed found to be abundant across the peptides and proteins examined in this study. The tracts differ considerably in size, depending especially on protein function. Together with the role of large same-charge zippers in function of histones and other DNA-interacting proteins [1, 6, 27], mRNA-binding proteins [16, 55], ribosomal proteins [3, 34, 41, 52] and shuttle proteins [15, 23], this invited characterization of size and ionic density as factors that could predict the frequency, affinity and stability of contacts with partners.

Methods

Definition of tracts

Homoionic tracts (also referred to as segments or zippers) were taken as continuous sequence segments starting and ending with ionizable residues of the same type of potential charge in the physiological [H^+] range (Asp and Glu for the homoacidic, His, Lys and Arg for the homobasic segments) and starting and ending by such residues. These segments are delineated by opposite-charged residues on either side. The exclusively acidic or basic motifs (e.g. *b*x**AA**x*b* and *a*x**BBBB**x*a*, where **A** = acidic, **B** = basic sidechain within the homoionic **AA** and **BBBB** tract, x = none or any number of non-ionic sidechains, and *b* and *a* are the delineating counter-charged residues) were surveyed separately from non-exclusive e.g. *b*x**A**x**A**x**A**x*b* and *a*x**B**x**B**x**B**x**B**x*a* tracts, which have the homoionic sections **A**x**A**x**A** and **B**x**B**x**B**x**B**. It is important to note that residues preceding and abutting the ionic termini of homoionic tracts are the opposite-charge ionic sidechains in at least 40 % of the tracts (data not shown), indicating a large potential for uses such as switching.

Polypeptide sets

Five sets of proteins and peptides were examined. These sets represent functionally very diverse categories of polypeptide. Sequences in the protein set (*pset*) include the broadest taxonomic ranges available in the reviewed records of the UniProtein and NLM/Protein databases. This set (listed in Table S1) contains 93 groups of proteins, with inclusion of orthologs in major taxonomic classes and of major subtypes. The selected polypeptidic agonists/hormones include only human sequences (*ago* set, Table S2), since few non-human peptidic agonists are fully characterized at present. Polypeptide/protein antagonists of receptor activity include human as well as non-human sequences (*ant* set, Table S3). Peptide/protein toxins from many taxa (*tox* set) are listed in Table S4. Antimicrobial peptides from various taxa, frequently called defensins, constitute *def* set (Table S5). It should be noted that, while effort was invested to collect representative sequences, no attempt was made toward exhaustive (database type) coverage of any particular protein or peptide type. All supplementary tables include access codes, numbers of residues in polypeptides, percentages of acidic and basic residues and cysteine, the percent sequence in the acidic and basic segments with two or more ionic residues, and the percent of sequence acidic and basic residues that is found in homoionic segments.

The *pset* groups are of two types. The first type includes groups of different proteins linked to the same broad purpose, such as ribosomal proteins, or membrane receptors with similar general role (e.g. integrins or opsins). For several groups (e.g. ribosomal proteins, linker proteins, shuttle proteins) this grouping entails considerable differences in size and structure. The groups of second type include functionally and structurally similar proteins, with the broadest available representation across taxa. Groups of this type include orthologs and close homologs of the respective proteins (such as hemoglobins $\alpha 1 \ldots 4$, $\beta 1 \ldots 3$, γ, δ, ε) across taxonomic groups. The above selections are supported by very similar parameter means (Table 1) between the entire protein set (4,384 sequences, including orthologs across taxa) and the human sequences in the set (2,324 primary orthologs only).

Table 1 Homoionic tracts and ionic constituents in the sets examined

Set and number of sequences	Residues per sequence	DE % in sequence	HKR % in sequence	Homoacidic % sequence	Homobasic % sequence	Homoionic % sequence	Homoacidic % all acidic	Homobasic % all basic	% HKR/ %DE	Homobasic/ homoacidic
pset all [4384]	504 ± 11.3	11.6 ± 0.06	14.7 ± 0.1	23.1 ± 0.43	34.4 ± 0.76	57.5	68.6 ± 1.3	78.9 ± 1.5	1.27	1.49
pset human [2323]	655 ± 20	11.6 ± 0.08	14.2 ± 0.09	23.2 ± 0.18	34.4 ± 0.24	57.6	69.7 ± 0.28	79.5 ± 0.19	1.22	1.48
pset other [2061]	334 ± 6.0	11.5 ± 0.10	15.2 ± 0.17	23.1 ± 0.26	34.4 ± 0.35	57.5	67.5 ± 0.39	78.2 ± 0.30	1.31	1.49
ago [431]	156 ± 9.5	9.9 ± 0.27	15.7 ± 0.26	16.7 ± 2.2	36.2 ± 4	52.9	52.1 ± 5.3	76.4 ± 7.4	1.59	2.17
ant [160]	299 ± 22.2	12.1 ± 0.32	14.6 ± 0.25	24.1 ± 3.3	31.8 ± 4.1	55.9	70 ± 8.1	78.6 ± 8.6	1.21	1.32
tox [851]	97.3 ± 6.3	9.34 ± 0.21	15.2 ± 0.25	14.2 ± 1.7	34.5 ± 3.5	48.7	46 ± 4.1	71.3 ± 6.1	1.63	2.43
def [704]	53.3 ± 0.72	8.32 ± 0.17	18.5 ± 0.22	11.8 ± 1.2	48.3 ± 4.6	60.1	45.5 ± 4.1	86.5 ± 7.2	2.22	4.03

All data are averages of parameters in the respective sets (Tables S1–S5), shown ±1 SEM

The column headings Set and number of sequences: the respective polypeptide set, followed by the number of sequences in the set (in brackets). Percent coefficients of variation for these are shown in Table 4. Residues per sequence: the average of amino acid residues per sequence in the respective group; DE % in sequence, HKR % in sequence: acidic and basic aa counts as % of total sequence aa count; Homoacidic % sequence, Homobasic % sequence: the respective homoionic aa counts as percent of total sequence aa counts; Homoionic % sequence: the sum of the preceding two columns; Homoacidic % all, Homobasic % all: the respective counts of the ionizable residues in homoionic tracts as percent of total counts of the ionizable residues in sequences; %HKR/% DE : (HKR % in sequence)/(DE % in sequence); Homobasic/homoacidic: (homobasic % sequence)/(homoacidic % sequence)

The surveyed complete sequences were extracted from UniProtein (http://www.uniprot.org) and Entrez/National Library of Medicine (NLM) (http:/ncbi.nlm.nih.gov./protein) websites; these sequences are identified in Tables S1–S6. Sequences of domains and motifs were extracted from Prosite (http://prosite.expasy.org) database, or from UniProtein and NLM (protein or nucleotide) records.

Data evaluation

The assessment of homoionic tracts was performed using Microsoft Excel macros. For helicity renderings, Protein Data Bank (PDB; http://www.rcsb.org/pdb/) data were used where available. Helicity predictions were generated in *porter* program [69] (http:/distill.ucd.i.e/porter). Sequence alignments were done in *blastp* program (based on [2]; http:/blast.ncbi.nlm.nih.gov./Blast.cgi), or with SSEARCH3 program [67]. Translation of RNA codes to protein sequence equivalents was done by an adaptation of FAATRAN program [66]. Regression and correlation tests were done in ProStat program (Poly Software, New York, NY, USA).

Surveys

Homoionic segregation is prominent and ubiquitous in proteins as well as in ligand-type peptides

The definition of homoionic tracts is outlined in the "Methods" section.

The distribution of ionic residues in the protein set (*pset*, Table S1) will be considered first, as its general pattern is also present in other sets. As seen in Figs. 1 and 2, the distribution has a low dependence on protein size above 100 amino acid residues (further abbreviated as aa) per chain. This is found for any of the parameters examined, including the bulk abundance of ionic residues in the sequence (lines 1 and 2 in Fig. 1), percent sequence in homoionic segments (lines 3 and 4 in Fig. 1; see also Fig. 2a) and percent of sequence ionic residues located in homoionic segments (lines 5 and 6 in Fig. 1) In *pset*, about 57 % sequence is in homoionic tracts that have two or more sidechains ionizable in the physiological [H+] range (Table 1). Most of the size-related difference in ionic content is found for sequences with less than 100 aa, mainly the highly basic polynucleotide partners (protamines, histones and short ribosomal proteins). The homoionic segments hold more than 68 % of all physiologically ionizable sidechains of *pset* (lines 5 and 6, Fig. 1; Table 1). In *pset*, homoionic tracts hold up to 79 % of all ionizable residues (Fig. 1 and Table 1), and occupy 40–70 % sequence in more than 80 % proteins (Fig. 3a). The collective fraction of sequence in homoionic tracts does not

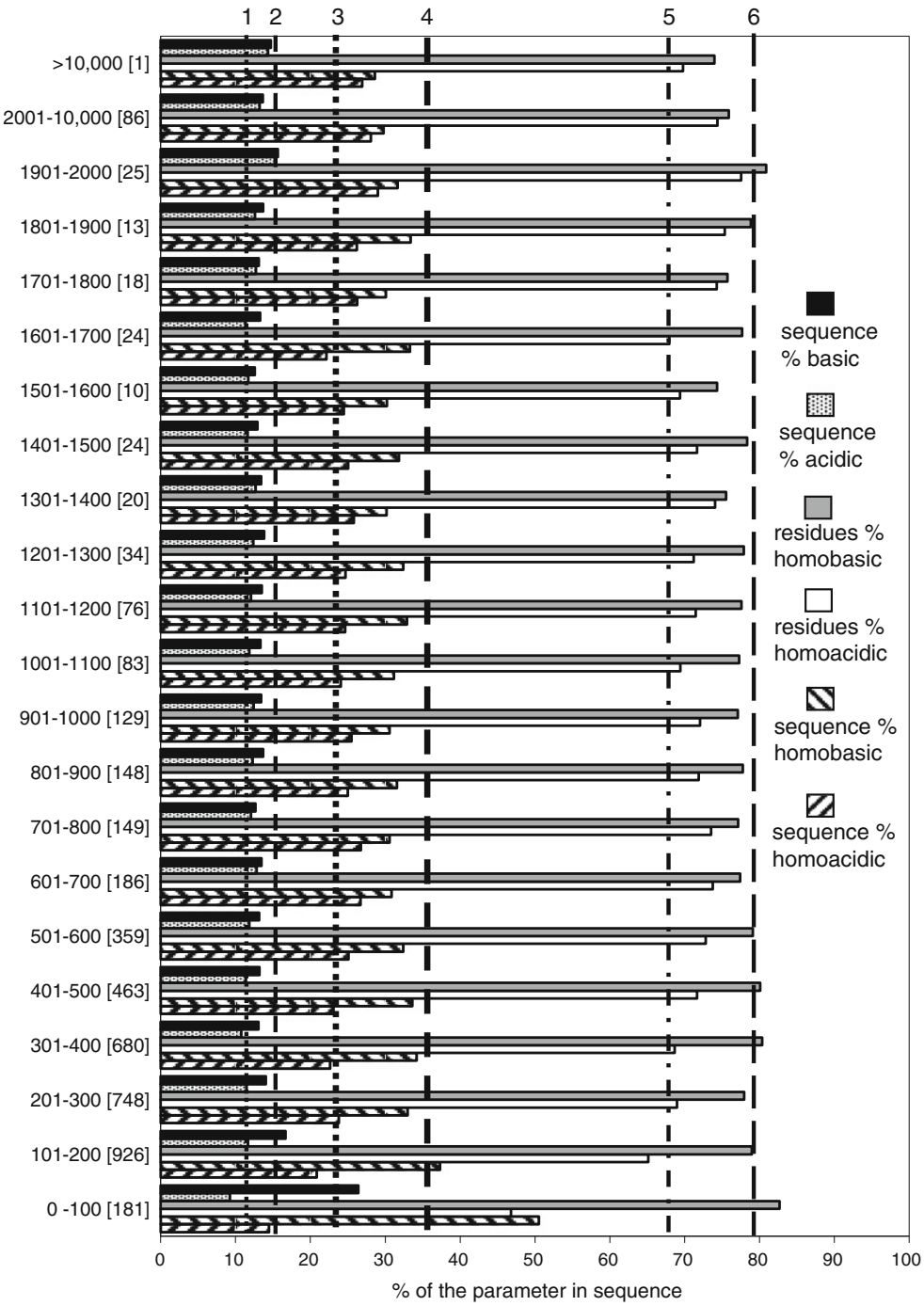

Fig. 1 Segregation of ionizable residues in the protein set examined (*pset*; Table S1) generally does not depend on protein chain length. The numbered lines from left to right are the means of: 1 acidic residues as % sequence amino acid (aa) count, 2 basic residues as % sequence aa, 3 % sequence aa in homoacidic segments with two or more ionic sidechains, 4 % sequence aa in the corresponding homobasic segments, 5 % of sequence acidic residues in ≥2-ionic homoacidic tracts, 6 % of sequence basic residues in ≥2-ionic homobasic tracts

vary radically with protein size (Figs. 1, 2). This fraction, however, does importantly depend on protein type (see Fig. 5), and varies extensively across individual proteins (Table S1).

It is important to note that the human sequences in *pset*, which include only single orthologs and no variant forms, show essentially no difference in homoionicity with the much more diverse non-human sequences (Table 1). However, there is an appreciable difference in the average chain length, due to scarcity of reviewed data on very large proteins (such as titin and nesprins) in other species. The bulk content of basic residues in the human part of *pset* is about 7 % lower than in the non-human part, due to the same cause (as can be deduced from Table S1), but this does not significantly affect the homoionicity parameters (Table 1).

The homoionic tract termini are preceded and followed by opposite-ionic sidechains in >40 % of the tracts, and by

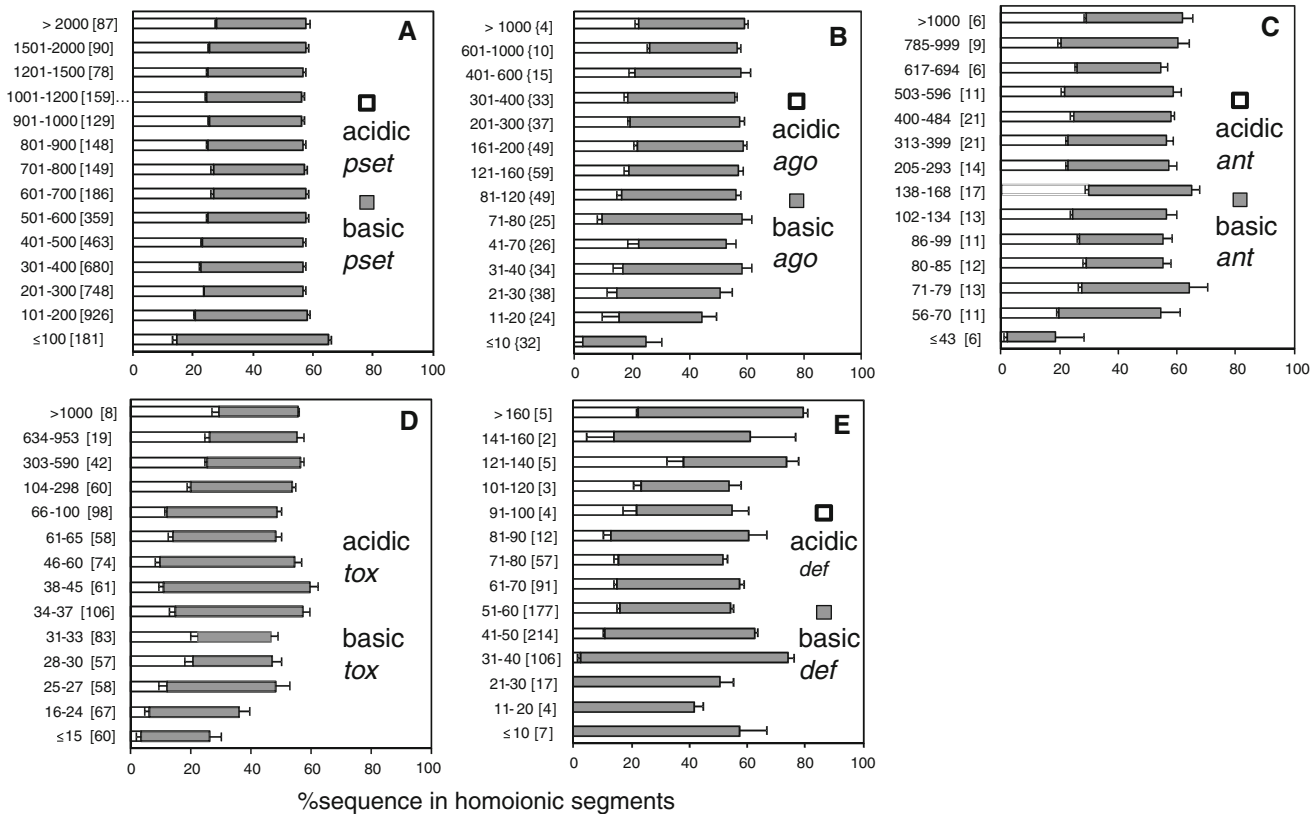

Fig. 2 Chain length and fraction of homoionic tracts in the sets examined. The proteins/peptides within sets were grouped by chain length as shown in the respective inscriptions. The number of sequences per group is shown in *brackets* following the *labels*. Data are the average percentages of sequence residues in tracts with two or more acidic or basic residues, shown with the corresponding standard errors. **a** The protein set (*pset*); **b** agonists (*ago*); **c** antagonists (*ant*); **d** toxins (*tox*); **e** defensins (*def*). For access and percentage data, see Tables S1–S5

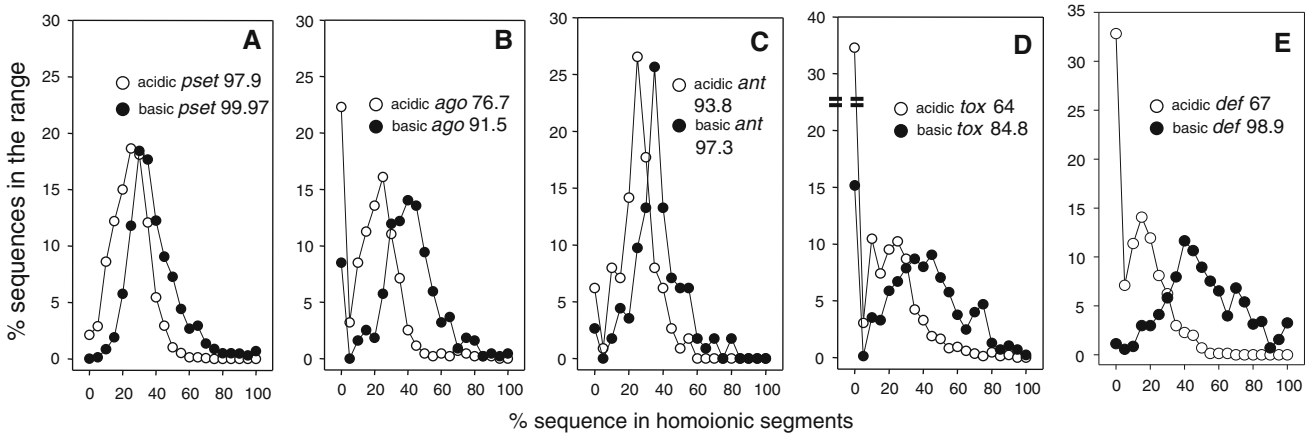

Fig. 3 Profiles of abundance of homoionic tracts in proteins and polypeptides examined. Shown are averages over 5 % increments of sequence length occupied by homoacidic and homobasic tracts with two or more ionizable residues for all molecules in Tables S1–S5. **a** The protein set (*pset*; Table S1); **b** agonists (*ago*; Table S2); **c** antagonists (*ant*; Table S3); **d** toxins (*tox*; Table S4); **e** defensins (*def*; Table S5). *Insets* show percentages of sequences that have homoacidic and homobasic tracts with two or more ionic sidechains

one or more non-ionic sidechains in >50 %. These neighbors of course are highly likely to affect interactions of the homoionic tracts, e.g. by switching reactivities in the case of counter-ionic residues, and by structural hinging and rotational activity in the case of proline [50, 84]. These parameters will be examined in a separate study.

The segregation does not importantly depend on chain length of the polypeptides

Excepting the shortest peptides, the collective fraction of sequence in homoionic tracts is not importantly related to size of the sequences (Figs. 1, 2). This fraction is significantly lower (just 40 %) in cysteine-rich conotoxin peptides (23 % Cys; Table S4). However, defensins, another group of cysteine-rich <100-aa peptides (14 % Cys; Table S5), show a segregation similar to >100-residue sequences of toxins, and nine *pset* proteins with more than 10 % bulk cysteine average 57 % sequence in homoionic tracts. The shortest members in all sets have almost exclusively basic homoionic tracts (Figs. 1, 2, 3; Tables S1–S5), with a large preponderance of basic residues in ligand-type peptide sets (Tables S2–S5 and graphs B–E in Fig. 3; see also Fig. 4). In *ago*, *ant* and *tox* sets, the shortest chains also show less than 50 % homoionicity (Fig. 2b–d). However, these chains represent small fractions of polypeptides in these sets (see Fig. 6). The *pset* proteins with ≤100 residues have a larger homoionicity than longer *pset* chains (Fig. 2a), and proteins with less than 50 aa have very large basic homoionicities (as can be deduced from Table S1). Excepting the shortest chains, the overall homoionic fraction of sequence is quite uniform relative to chain length in all sets, and especially in *pset* (Figs. 1, 2).

Acidic sidechains consistently show lower segregation than the basic

Inclusion of basic residues in homoionic tracts is consistently higher than that of acidic for all sets, and especially for toxins and defensins (Table 1; Figs. 2d, e, 3d, e). In *pset*, almost all sequences have homobasic tracts (Table S1; Fig. 3a). In ligand-type polypeptides (the *ago*, *tox* and *def* sets and part of the *ant* set), a lower degree of homoacidic sequestration accompanies the low general abundance of acidic residues. This should help electrostatic matching with anionic segments in receptor sequences (e.g. those in the N-terminal extracellular domain of GPCRs; Table 6).

Assuming a random representation, the three basic residues would outnumber the two acidic by 1.5 in the entire sequences. As seen in Table 1, the %HKR/%DE bulk ratio is quite below the random for *pset* (1.27) and *ant* (1.21), is somewhat above the random in *ago* (1.59) and *tox* (1.63), and is considerably basic-biased in *def* (2.22). However, the ratio of the fraction of sequence in homobasic to that in homoacidic tracts is above 2 for ligand-type sets (and even above 4 for *def*), and also exceeds the bulk ionic sidechain ratio in *pset* (1.49 vs. 1.27) and *ant* (1.32 vs. 1.21). The homobasic bias for ligand-type peptides should connect to frequent quite anionic contact motifs of partners, and also to somewhat anionic overall charge of the bilayer.

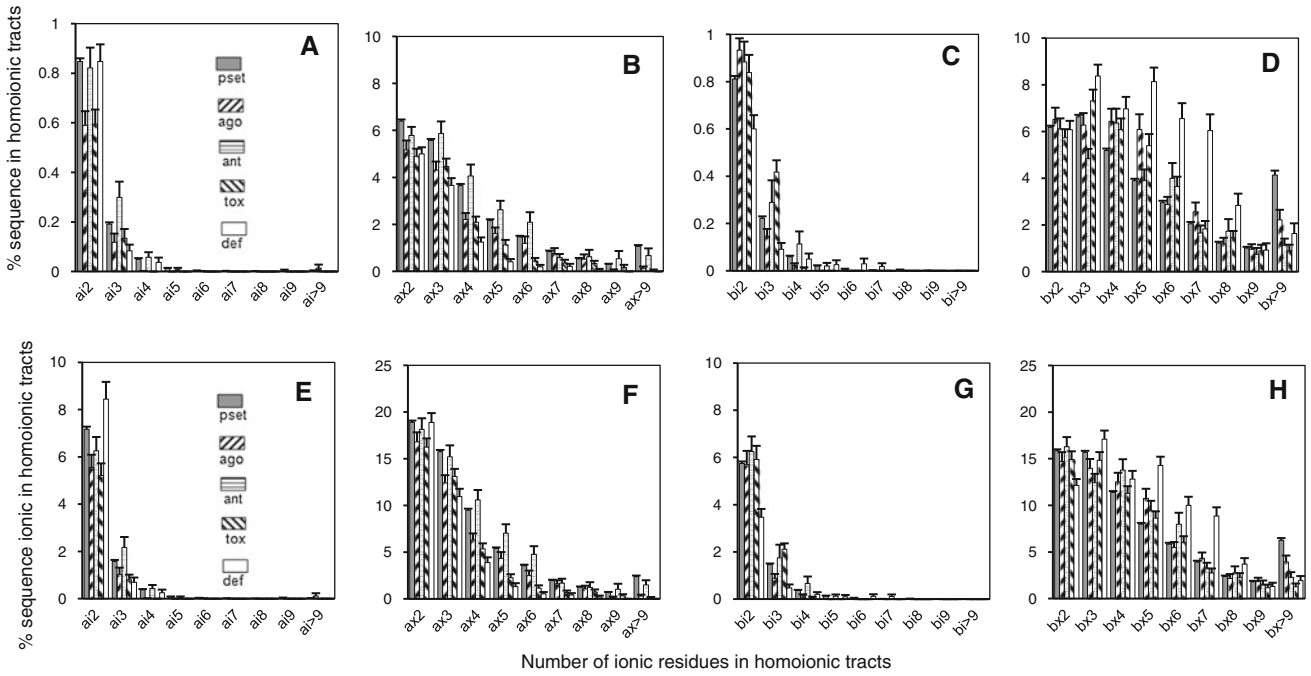

Fig. 4 Percent sequence residues in homoionic tracts with two or more ionizable sidechains. This is presented separately for tracts containing only ionic residues and for tracts not exclusively ionic (see "Methods"). *Upper row* fractions of sequence length occupied by ≥2-homoionic tracts. **a** Ionic-only acidic (ai); **b** non-exclusive acidic (ax); **c** ionic-only basic (bi); **d** non-exclusive basic (bx). *Lower row* fractions of sequence ionic residues found in ≥2-homoionic tracts. **e** Ionic-only acidic (ai); **f** non-exclusive acidic (ax); **g** ionic-only basic (bi); **h** non-exclusive basic (bx)

Isolated homoionic doublets are frequent in all sets and may have functional distinction

Isolated acidic or basic pairs (*b*x**AA**x*b* and *a*x**BB**x*a*; see "Methods") are far more abundant than the longer ionic-only homoionic segments in all sets examined (Fig. 4). In *pset*, the isolated homo-2-ionic segments contain 4.3-fold more acidic and 3.6-fold more basic residues than the corresponding tri-ionic segments. This would add weight to consideration of these pairs as functionally distinct from non-exclusive 2-homoionic (*b*x**A**x**A**x*b* and *a*x**B**x**B**x*a*) tracts. Similar could apply to segments with larger numbers of exclusively ionic sidechains, which, however, are much less represented than the segments not exclusively ionic (Fig. 4). The segregated ionic-only segments represent in all sets only a minor part of sequence length (Fig. 4a, c), but a sizable fraction of sequence ionic residues (Fig. 4e, g). This is especially prominent for 2-acidic motifs, particularly in defensins (Fig. 4e).

Reports about functional impact of homoionic multiplets in polypeptides do not distinguish the isolated, counter-ion segregated motifs from quite numerous homoionic multiplets (usually doublets) found within larger homoionic sequence segments (as in CXCR4 [45], FGF [75] and Tat-HIV [18]). In the AT1A receptor (u-p P30556), an isolated DD pair is important for supporting phosphorylation [57]. Diacidic motifs could generally help kinase activity [19, 29]. These examples support especially a docking role for diacidic isolates. However, an isolated DE pair is important in arrangement of helices in the *trp* operon aporepressor [46] of many enterobacteria, and diacidic pairs regulate activity of several DNA-handling enzymes.

Dibasic motifs rarely have been studied in isolation, due to their frequent repetitive occurrence in larger basic zippers in many proteins. "KR-rich" regions attracted much interest (e.g. [94]), but they are frequently parts of larger homobasic tracts. However, dibasic isolates are frequent especially in nuclear localization motifs (see e.g. [22]). Basic doublets within non-exclusive basic zippers could form ligand binding sites, e.g. for PIP2 [13].

Defensins have the largest fraction of sequence and of sequence basic residues in 3 to 8-ionic basic zippers (Fig. 4d, h). The large presence of >9-basic zippers in *pset* (Fig. 4d, h) is found not only for polynucleotide-handling proteins, but also for heptahelical GPCRs (Fig. 5b).

Multi-ionic homoionic tracts

Homoionic tracts with more than five ionic residues are found especially in proteins engaging in protracted or even stable interactions with protein or polynucleotide partners, and in ligand peptides that bind to partners having similar opposite-charge homoionic segments. Such zippers collectively contain more than 10 % sequence residues in any of the five sets, and more than 30 % ionic sidechains in all sets except *tox* (Table 2). Ratios of sequence fractions in basic and acidic >5-ionic tracts are not very high in *pset*, *ago* and *ant*, but rise over 6 in *tox* and over 30 in *def*, indicating much higher relative frequencies of multibasic tracts in the latter two sets. The sequence basic/acidic residue ratio for >5-tracts is, however, similar for the first four sets, due to larger densities of acidic sidechains in the corresponding tracts (see the next section), but is still very high in *def* (which miss ≥9-homoacidic tracts, Fig. 7f).

Figure 5 shows percentages of >5-homoionic tracts for 20 groups of *pset* proteins with the highest and the lowest respective fractions of sequence. The most homoacidic groups (Fig. 5a) include, as could be expected, several groups of linking, platform, docking, shuttling and chaperone proteins, while there is just one group of enzyme proteins. The presence of transcription factors from many taxa could reflect use in separation of DNA-engaging proteins. The most basic scores (Fig. 5b) include, in addition to protamines and histones, ribosomal proteins from all taxa. This could attest to validity of the multi-homoionic criterion in predicting stable association with polynucleotides. Large homobasic tracts are not prominent in non-ribosomal RNA-binding proteins (which frequently form dynamic complexes with mRNAs). The visual and non-visual heptahelical A-GPCRs are also in the list, due to strongly basic intracellular parts (see Table 6), and human RNAses are the only enzymes in the most basic >5-homoionic groups.

The low-scoring acidic groups (Fig. 5c) in addition to DNA-targeting protamines and histones predictably feature three groups of RNA-binding proteins. The 14.3.x platform proteins quite lack the multi-basic zippers (while being strong on multi-acidic, Fig. 5a). This lack is shared by three GTPase transducer groups, opsins, and five groups of heme-binding proteins (Fig. 5c). The low multi-basic finds (Fig. 5d) include several highly multi-acidic groups (Fig. 5a). Interestingly, the RNA-binding *pumilio* repeat proteins are also low in multi-homobasic segments, which could be reflected in their complex association with UGUANAUA mRNAs [88]. The archaeal opsins which transduce to membrane-embedded proteins (e.g. 44) are devoid of both acidic and basic >5-zippers (Fig. 5c, d).

The largest number of ionic sidechains per homoionic tract in the five sets is generally not much above 10 (Table 3). Tracts with 10 or more ionic sidechains are found in, respectively, 11.9 % (acidic) and 21.1 % (basic) sequences in *pset* (Table 3), and *ant* set also shows numerous large homoionic zippers. In *pset*, multiple >9-ionic zippers are found especially in eukaryotic 60S ribosomal proteins and basic Leu zipper proteins, and also in long linkers; in titin, 12 acidic and 4 basic >9-zippers could have use in

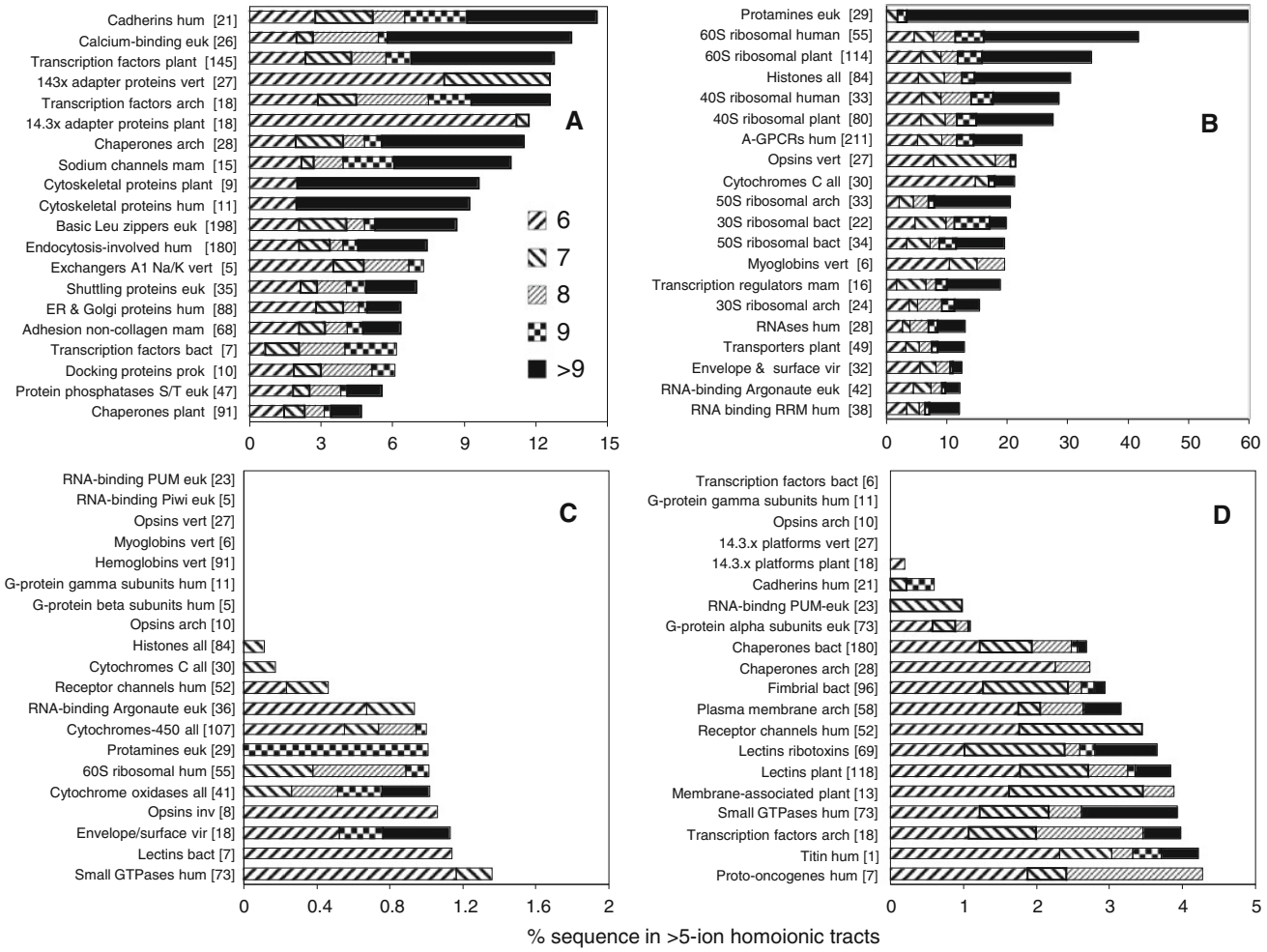

Fig. 5 Groups of *pset* proteins with largest and lowest abundance of >5-ionic homoionic tracts. *Upper row* 20 groups with highest % of sequence residues in **a** >5-ion homoacidic and **b** >5-ion homobasic tracts. *Lower row* 20 groups with lowest % sequence in **c** >5-ion homoacidic, **d** >5-ion homobasic tracts. For group constituents, see Table S1. *prok* prokaryotes, *euk* eukaryotes, *hum* human, *bact* bacterial, *arch* archaeal

Table 2 Fractions of sequence residues and ionic residues in >5-homoionic tracts

Group	% Sequence in >5-acidic tracts	% Sequence in >5-basic tracts	(>5-basic)/(>5-acidic) for % sequence*	% All acidic aa in >5-acidic tracts	% All basic aa in >5-basic tracts	(>5-basic)/(>5-acidic) for residue fraction[a]
pset	4.16 ± 0.22	11.1 ± 0.45	2.66	9.75 ± 0.5	18 ± 0.73	1.85
ago	2.58 ± 0.83	8.42 ± 1.7	3.26	8.49 ± 1.6	22.8 ± 2.8	2.68
ant	4.53 ± 1.5	9.26 ± 2.1	2.04	10.4 ± 3	23.1 ± 3.7	2.22
tox	1.4 ± 0.52	8.79 ± 1.6	6.28	6.42 ± 1.0	14.1 ± 2.3	2.20
def	0.48 ± 0.23	18 ± 2.6	37.5	3.22 ± 0.53	28.5 ± 3.5	8.85

The data are sums of the respective percentages in ionic-only and non-exclusively ionic tracts. See Fig. 4 for discrete values for the ionic-only tracts and for the tracts not exclusively ionic

[a] Ratio of the basic to acidic parameter in the preceding two columns

anchoring [97]. There are very few cases of exclusively ionic >9-homoionic tracts (in *pset*, five of 704 > 9-acidic and just one of 1,134 > 9-basic segments).

Homobasic >9-ionic tracts are found in about 10 % of agonist sequences (Table 3), which could be connected especially to chemotactic mobility. Only about 2 % of

Table 3 Parameters of >9-ionic homoionic tracts in the sets examined

Set [number of sequences]	Acidic aa per tract	% Acidic density	Found in [%][a]	Basic aa per tract	% Basic density	Found in [%][a]
pset [4384]	13.1 ± 0.2	39.8 ± 0.75	522 [11.9]	13.0 ± 0.15	33.6 ± 0.5	923 [21.1]
ago [431]	12.3 ± 0.71	30.9 ± 1.7	6 [1.39]	11.8 ± 0.39	38.5 ± 2.5	43 [9.55]
ant [160]	12.6 ± 1.1	37.1 ± 3.3	11 [6.88]	11.5 ± 0.54	32.5 ± 2.5	18 [11.3]
tox [851]	11.0 ± 0.60	28.2 ± 0.48	7 [0.82]	10.7 ± 0.50	34.8 ± 3.4	16 [1.88]
def [704]	–	–	–	11.3 ± 0.35	32.6 ± 2.3	15 [2.13]

[a] The number of set's sequences that have >9-ionic zippers, followed in the brackets by that number as % of all sequences in the set

toxins and defensins have such segments (Table 3), in line with their chiefly insertive mechanisms of action [40, 64].

Ionic density within the sets does not depend appreciably on the polypeptide chain length

An evaluation of the density of ionizable residues in non-exclusive homoionic tracts of polypeptides in different size ranges is presented in Fig. 6; this complements the analysis of homoionic segments presented in Fig. 2. It should be noted that the density variation within the entire sets (regardless of size ranges) is below 33 % for *pset*, *ago* and *ant* sets (Table 4), which is in contrast with large variation in chain length of the respective polypeptides. This indicates a fair degree of homogeneity in this density across the diverse polypeptide species included. The variation is, however, above 40 % for acidic *tox*, and above 35 % for basic *tox* and *def* densities (Table 4).

With *pset* (Fig. 6a), the overall ionic densities are close to 33 % for both acidic and basic tracts (Table 4). The ≤100-aa proteins show a higher density for both acidic and basic tracts compared to larger chains (Fig. 6a). Chains with 300 or more residues show quite uniform average densities. The agonists (Fig. 6b) average 36.5 % acidic and 34.9 % basic density (Table 4), with basic tracts having a higher density than acidic in <71 peptides, and with the acidic density higher than the basic for longer chains. The *ant* sequences (Fig. 6c) have 39.9 % acidic and 36.5 % basic mean density (Table 4), and the acidic density is higher than the basic for all size ranges (Fig. 6c). The same is observed with *tox* set (Fig. 6d), which has the largest mean density among the sets for both acidic (40.5 %) and basic (35.8 %) homoionic tracts (Table 4). The *def* set (Fig. 6e) has mean 37.5 % acidic and 35.6 % basic homoionic density (Table 4) and consistently shows larger acidic densities above the chain length of 30 (Fig. 6e).

As is clear from Fig. 6, the ionic density within homoionic tracts does not have a large dependence on chain size for polypeptides in the size ranges that have strong representation in the respective sets. Large variations are chiefly associated with polypeptides representing <10 % of the respective sequences.

The ionic density generally changes little with size of the homoionic tracts

The size of homoionic segments shows strict correlation (in Pearson and Kendall tests) with the number of ionic sidechains for any of the sets. The linear regressions are very significant in all cases (Fig. 7). Slopes of the regressions (Fig. 7a) range from 2.98 in *ant* to 3.45 in *def* for acidic, and from 3.49 in *ago* to 3.62 in *pset*, indicating addition of about 3.3 total residues per acidic, and about 3.5 per basic sidechain. The *Y* axis intercepts for acidic segments range from −1.42 % in *def* to 0.306 % in *ant*, and for basic tracts from −1.58 % in *def* to −1.1 % in *ago* set (Fig. 7b), which strongly supports linearity of changes. It should be noted that both the average densities and the regressions for 2–6 ionic segments (i.e. for homoionic tracts that are strongly represented in all sets; Fig. 4) are much closer across the sets (Fig. 7d, e), and also are free from deviations seen in poorly represented >6-ionic (especially >6-acidic) tracts of ligand-type sets. In accordance with the density data, the number of homoionic tracts per 100 sequence residues for acidic is 1.44 for *def* and 2.06–2.41 in other sets (Fig. 7c), reflecting on both the generally lower segregation of acidic sidechains, and in particular on the low bulk acidity of defensins (Table 1). For the homobasic tracts, this number across the sets is larger as well as closer, from 2.79 in *pset* to 3.43 in *def* (Fig. 7c). The distribution of densities across tracts versus increase in the number of ionic sidechains also indicates lack of major change with ion number (Fig. 7f, g). Some deviations are, however, seen with >6-acidic segments especially in short defensin molecules (Fig. 7f).

The >9-homoionic tracts in *pset* generally do not have large ion densities (Table 3). Protamines present an exception, with density of 70 % in basic >9-zippers (and 65 % for all basic zippers). In 395 ribosomal proteins, 258 > 9-basic tracts average 39 % density, while the transductionally dynamic A-GPCRs average the density of only 23 % in 105 > 9-basic zippers of 210 human receptors.

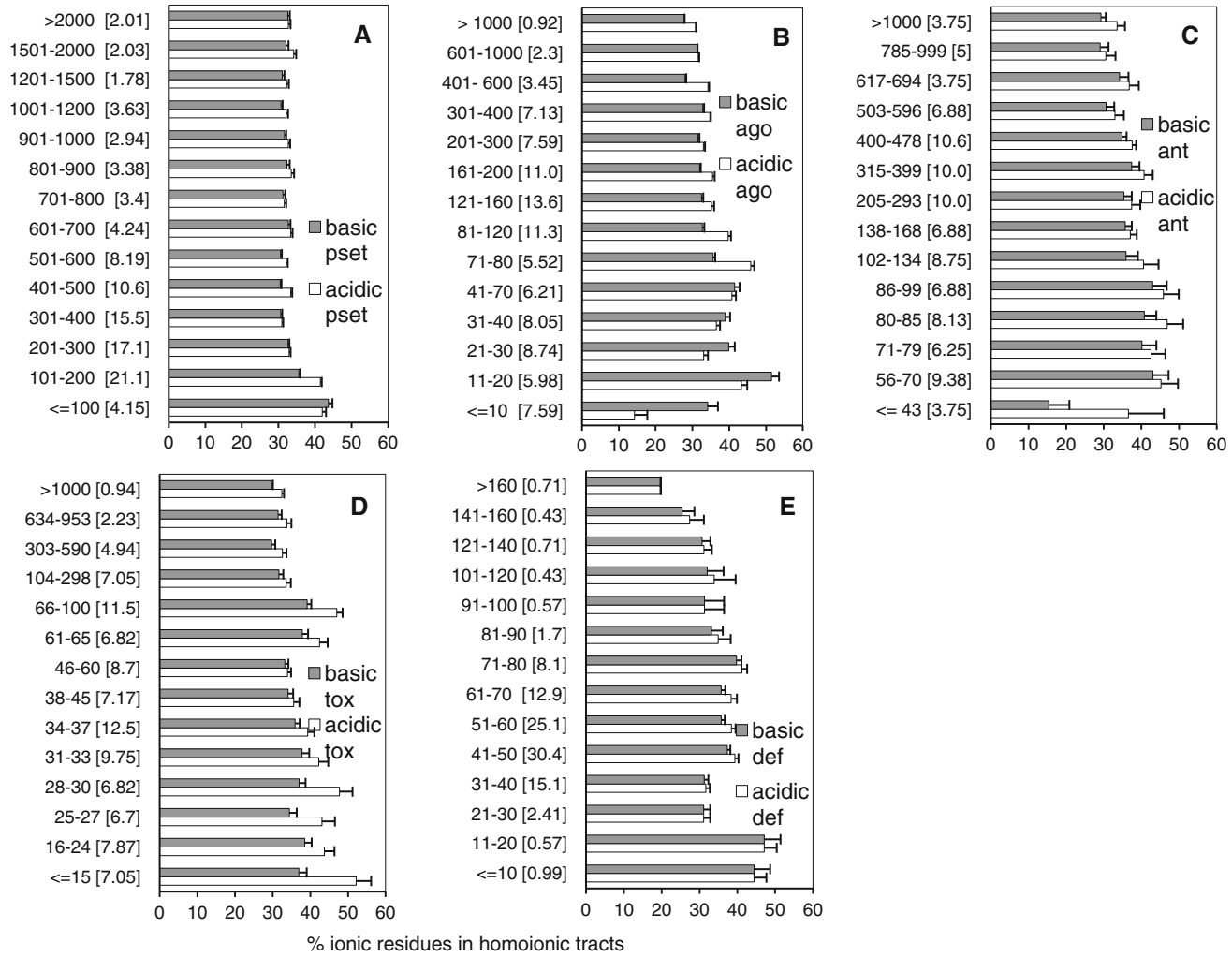

Fig. 6 Chain length and density of ionic residues in homoionic tracts across the sets examined. Percentages of the set sequences in the chain length ranges are shown in *brackets*. For numbers of sequences in the ranges, see Fig. 2

Table 4 Parameters of ionic residue density in non-exclusive homoionic tracts of the sets

Set [# sequences]	aa per sequence [% cv]	% Acidic density [% cv]	% With homoacidic tracts[a]	% Basic density [% cv]	% With homobasic tracts[a]
pset [4384]	503.7 [148]	33.8 [29.6]	97.9	32.9 [29.2]	99.7
ago [431]	156.4 [79.2]	36.5 [29.9]	73.6	34.9 [32.4]	88.9
ant [160]	298.7 [93.9]	39.9 [31.8]	98.1	36.5 [28.5]	97.5
tox [851]	97.25 [185]	40.5 [45.2]	84.8	35.8 [33.9]	77.4
def [704]	53.26 [35.3]	37.5 [36.7]	98.9	35.6 [32.6]	98.0

[a] Note that these values refer to tracts not exclusively ionic. The parameter means are shown with percent coefficients of variation (100*standard deviation/mean) in brackets. Values for all homoionic tracts are shown in inscriptions of Fig. 3 graphs

Inteins and intron-derived proteins are strongly homoionic and intronic mRNA sequences are rich in potentially homobasic code

The consistently large homoionic segregation in all polypeptide sets prompted an examination of possible equivalents in untranslated mRNAs and in non-canonical proteins. Introns represent a large part of eukaryotic genes, and in some cases even get translated into functional proteins [30, 78]. Inteins, the self-excising protein-splicing proteins [28, 32] are considered as protein equivalents of introns. Introns of eukaryotic transcripts share the

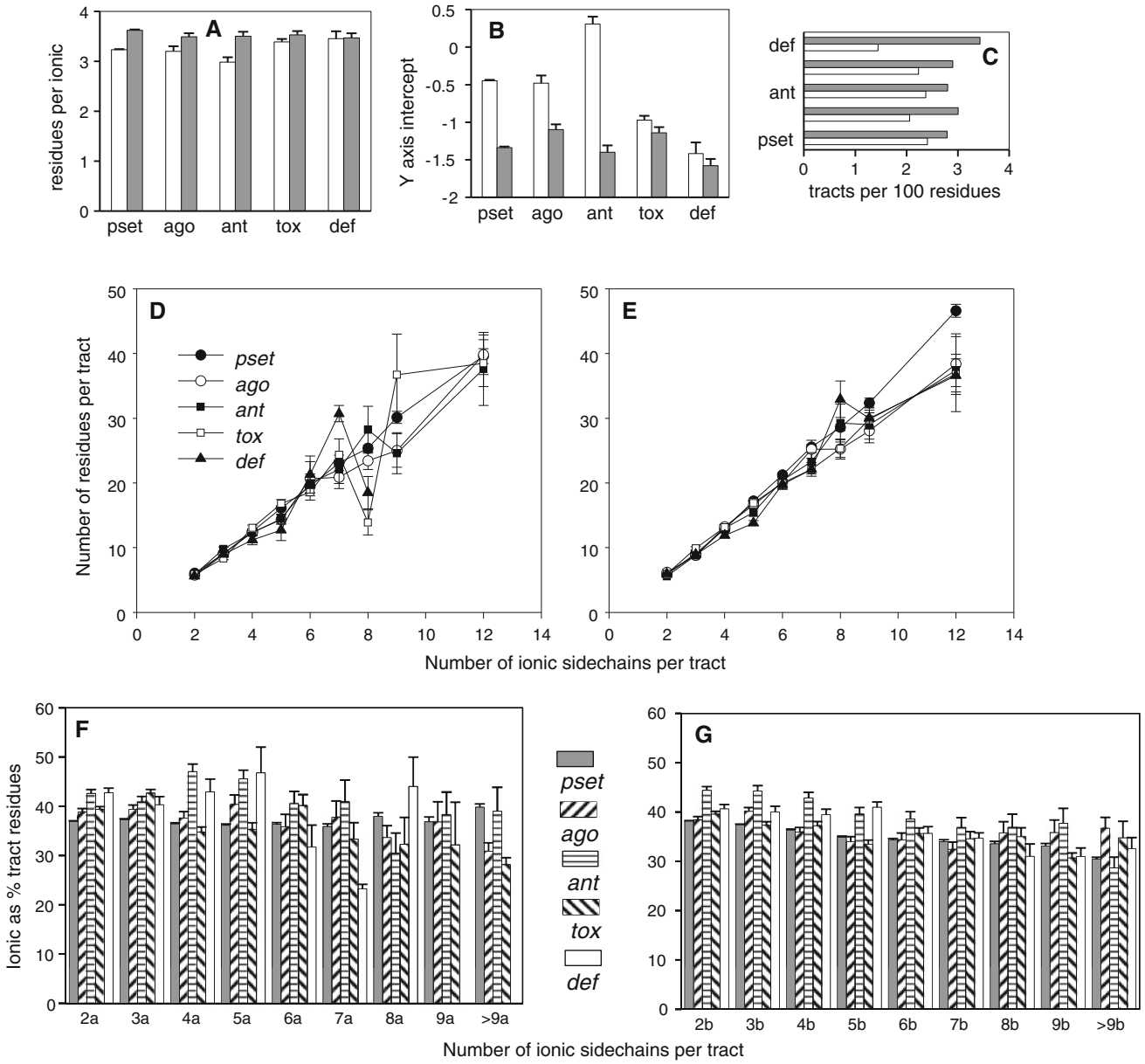

Fig. 7 Characterization of ionic density of homoionic non-exclusive segments with two or more ionic sidechains in the sets examined. **a** Slopes of linear fits (data shown in **d**, **e**) of numbers of tract residues versus number of ionic sidechains per tract; **b** *Y* axis intercepts of these fits; **c** numbers of homoionic tracts per 100 residues. In **a**–**c** acidic segment bars are blank. **d** Acidic homoionic tract size versus number of ionic sidechains; **e** the same for basic tracts; **f** % density of ionic sidechains in acidic tracts. The data are shown with the respective standard errors

nucleoprotein packaging and processing with exons [80], but may not affect the physiological translation. Regardless of the potential physiological impact of introns, it is of interest to compare exons and introns in terms of coding that does (with exons) or may (with introns) produce homoionicity.

A comparison focusing on nucleases as proteins functionally similar to endonucleases expressed from introns (Table 5) indicates that the hypothetical intronic transcripts, the physiological intron-derived endonucleases and the protein parts of bacterial RNAses-P also have a low acidic residue content and homoacidic portion of sequence. The 88 examined non-translated mRNA sequences from 20 genes (Table S6) include 79 declared introns and nine 3′-UTR, 5′-UTR and other untranslated transcript sections. More than 90 % intronic constructs show large homobasic bias (Table S6). The constructs would have only 7 % sequence in homoacidic zippers. This is followed by the third intracellular loop (ic3) of mammalian Y5 NPY receptor (which in itself could be an intron-derived, frameshift-induced insert [37]) at 7.5 %, RNAse-P protein parts at 10.3 % and intron-derived endonucleases at 13.6 %

Table 5 Homoionicity in intron-like proteins, nucleases and intronic constructs

Group	DE%	HKR%	Homoacidic % sequence	Homobasic % sequence	Homoacidic % all acidic	Homobasic % all basic	% HKR/ % DE	Homobasic/ homoacidic % sequence	Homobasic/ homoacidic % all
Restriction enzymes [29]	13.7 ± 0.31	14.3 ± 0.26	27.2 ± 0.95	29.3 ± 0.96	74.1 ± 1.3	77.1 ± 0.95	1.04	1.08	1.04
Inteins [12]	12.8 ± 1.3	13.8 ± 0.74	25.7 ± 2.4	27.1 ± 2.4	75.6 ± 4.4	72.8 ± 3.3	1.08	1.05	0.96
RNAses *E. coli* [14]	13.3 ± 0.66	15.3 ± 0.87	24.5 ± 2.4	31 ± 3.5	69.8 ± 2.4	75.7 ± 2.1	1.15	1.27	1.08
DNAses human [7]	11.2 ± 1.3	13.3 ± 1.4	23 ± 3.8	33.1 ± 2.9	69.1 ± 3	79.1 ± 4	1.19	1.44	1.14
DNAses *E. coli* [5]	11.5 ± 0.72	16.1 ± 1.1	18 ± 3.1	35.3 ± 2.8	64.5 ± 4.6	75.6 ± 4	1.40	1.96	1.17
RNAses human [28]	10.2 ± 0.59	15.3 ± 0.62	17.4 ± 1.8	41.3 ± 2.1	60.1 ± 4.6	84.3 ± 1.5	1.50	2.38	1.40
Endonucleases from introns [36]	8.27 ± 0.46	14.9 ± 0.63	13.6 ± 1.1	40.4 ± 1.27	54.8 ± 2.7	85.9 ± 1.2	1.80	2.97	1.57
RNAses-P *E. coli* [22]	7.49 ± 0.05	24.5 ± 0.08	10.3 ± 0.4	69.2 ± 0.57	63.1 ± 2.2	96.4 ± 0.14	3.27	6.72	1.53
Ic3 loop of NPY Y5 receptor [5]	8.26 ± 0.27	22.7 ± 0.37	7.52 ± 0.54	52.5 ± 2.0	57.3 ± 2.34	89.6 ± 1.28	2.75	6.98	1.56
Intronic constructs [88]	6.45 ± 0.32	13.8 ± 0.41	7.08 ± 0.75	35.2 ± 1.6	38.2 ± 2.9	77.7 ± 2.1	2.14	4.97	2.03
Expressed i.c.[a] proteins [20]	11.9 ± 1.5	16.6 ± 1.6	21.9 ± 3	41.0 ± 3.8	65.9 ± 3.6	79.9 ± 3.9	1.39	1.87	1.21

Homobasic/homoacidic % all: the ratio of fractions of sequence basic and acidic residues found in the respective homoionic tracts

For explanation of other column headers see the legend of Table 1. For individual data and access codes see Table S6

[a] Averages of parameters for 20 expressed proteins from genes bearing 88 introns that were used to construct the hypothetic intronic protein sequences (see Table S6 for the list)

homoacidic sequence fraction (Table 5). The homobasic sequence fraction in all of the above groups is higher than in inteins or the restriction enzymes. The bulk basic/acidic residue ratio is above 2 for intronic constructs, the Y5 ic3 loop, and RNAses-P, above 1.5 for human RNAses and intron-derived endonucleases, and 1–1.4 for all other groups (Table 5). The homobasic/homoacidic sequence fraction ratio is 5–7 in intronic constructs, the Y5 loop and RNAses-P, and less than 2 in all other groups including the exonic correspondents of the constructs (Table 5). The ratio of homobasic-located basic to homoacidic-situated acidic aa is 1.4 or above for RNAses, intronic endonucleases, RNAses-P and ic3 of the Y5 receptor, and above 2 for intron constructs (Table 5). The self-excising inteins have much higher homoacidic content than intron-derived endonucleases. Other nucleases in Table 5 have 18–27 % homoacidic sequence, and *pset* shows 23.1 % (Table 1). The above differences may relate to excision mechanisms of intronic mRNA sequences, and point to the potentially large homobasic bias as opposing translatability of introns.

Homoionicity should be important in selection of partners by many protein domains

Table 6 shows parameters of homoionicity in a number of protein domains with generally well-defined functions. These parameters are briefly considered below in connection to the functions.

Basic Leu zippers contain a long homobasic segment with a large ionic density (Table 6) that helps high-affinity binding to DNA (e.g. [43]). This tract is followed by a mainly helical leucine-rich segment mixing large-density homoacidic and short homobasic tracts. This type of composition would be expected for a domain that needs to accommodate twin functions of binding to DNA and tethering histone/repressor complexes (e.g. [93]). The strong homobasic bias (Table 6) is likely to, in addition to DNA binding, also help homing of acetylated repressors [17, 93].

The type I calcium-binding EF hand-1 domains essentially are homoacidic tracts and have large acidic and very

low basic density (Table 6). These parameters fit the requirements for tertiary structuring of EF hands [9, 33].

The caspase-handling CARD and Death domains are similar in bulk content of ionic residues as well as in homoionicity and ionic density (Table 6). The numerous short homoacidic tracts in these domains could serve in docking and subcellular relocation [48].

The annexin-type domains have large-density short homoacidic tracts which should be involved in calcium binding [90]. The more represented homobasic tracts should be chiefly responsible for binding to the bilayer phospholipid [72]. However, short basic zippers are often involved in type II and III calcium-binding sites.

The *src*-homology 2 (SH2) domains and *src*-homology 3 (SH3) domains show moderate homoionic densities (Table 6), forecasting mainly contact and adapter roles, as seen in adapter protein Grb2 [51], endocytosis-involved proteins (especially dynamins [12, 85]), cytoskeletal linkers, and many enzymes (e.g. protein kinases and phosphatases). The clearly stronger acidic zippers in SH3 and basic zippers in SH2 could indicate a "division of labor" for these well-represented domains, which frequently reside in the same protein.

The frequent pleckstrin homology (PH) domains should serve especially in dynamic adhesion to membrane phospholipids. The PH domains are strongly homobasic (Table 6), therefore preferring phosphoinositides [4] and phosphatidylserine [35].

The very high homobasic character and rather low homoacidity of intracellular tails (ic4 domains) of human A-GPCRs (Table 6) indicate preference for acidic counterparts in transducers, antagonists and effectors. This corresponds with strong homoacidic zippers in partners, including phospholipases [14], nucleotidyl cyclases [26] and protein kinases [71], and should also be helpful in association with G-protein heterotrimers in the ER [95]. The H8 helix of A-GPCRs at the N-terminus of the ic4 domain [59] is a homobasic tract in ~90 % of non-visual receptors and in nearly all eukaryote opsins (Table S4 in [62] and may relate to excision mechanisms of intronic mRNA sequences). The dynamic transduction of these receptors to Gα subunits could involve short

Table 6 Homoionicity in some widely represented protein domains

Group and number of domains	Database access	# aa in domain	% Domain's aa in acidic tracts	% Domain's aa in basic tracts	% Domain's DE in acidic tracts	% Domain's HKR in basic tracts	% Ionic in acidic tracts	% Ionic in basic tracts
Basic Leu zippers eukaryote [26]	PS50217	58.6 ± 1.1	17.6 ± 1.3	50.2 ± 2.0	63.2 ± 3.3	88.2 ± 1.3	57.4 ± 3.2	52.4 ± 1.7
EF hand-1 eukaryote [3657]	PS00018	13	57.6 ± 0.5	3.55 ± 0.18	76.7 ± 0.17	9.87 ± 0.47	49.0 ± 0.41	6.94 ± 0.34
CARD [76]	PS50209	86.8 ± 0.6	26.2 ± 1.14	27.3 ± 0.84	71.1 ± 1.7	77.2 ± 1.4	46.7 ± 2.1	47.0 ± 1.3
Death [174]	PS50017	79.5 ± 0.61	29.2 ± 0.85	23.3 ± 1.0	70.8 ± 1.0	64.0 ± 1.7	42.3 ± 1.0	46.3 ± 1.3
Annexin eukaryote [317]	PS00223	53	22.4 ± 0.2	28.2 ± 0.66	74.6 ± 1.2	72.0 ± 1.1	57.6 ± 0.5	42.6 ± 0.79
Src-homology 2 (SH2) [117]	PS50001	96.3 ± 0.74	15.9 ± 0.83	33.8 ± 1	58.5 ± 1.9	79.4 ± 1.1	42.6 ± 1.3	38.2 ± 0.85
Src-homology 3 (SH3) [267]	PS50002	61.7 ± 0.31	31.3 ± 0.93	21.3 ± 0.79	77.4 ± 1.1	69.4 ± 1.2	40.3 ± 0.8	42.2 ± 1.0
Pleckstrin homology (PH) [283]	PS50003	112 ± 1.5	18.1 ± 0.49	38.4 ± 0.69	65.3 ± 0.92	82.4 ± 0.63	41.4 ± 0.76	36.9 ± 0.51
Ic4 domain of A-GPCRs [210]	UniProtein[a]	56.5 ± 2.2	12.3 ± 0.88	44.1 ± 1.3	46.4 ± 2.5	84.7 ± 1.0	35.5 ± 2.2	38.8 ± 0.92
Ec1 domain of A-GPCRs [210]	UniProtein[a]	49.8 ± 3.9	28.5 ± 1.8	12.9 ± 1.2	65.1 ± 2.7	40.0 ± 2.9	31.7 ± 1.9	21.9 ± 1.9
Transmembrane domains of A-GPCRs [1470][b]	UniProtein[a]	23.4 ± 0.02	2.24 ± 0.26	5.96 ± 0.48	7.59 ± 0.69	12.8 ± 0.89	2.53 ± 0.25	7.16 ± 0.53

All percentages are averages for the entire domains. Sequences not identified with taxa are human

number of, *aa* amino acid residues, *Ic4* intracellular domain #4 (C-terminal in all GPCRs), *Ec1* extracellular domain 1 (N-terminal in all GPCRs)

[a] The access codes are listed in Table S1

[b] Averages for the seven transmembrane domains of 210 human A-GPCRs

homoacidic tracts present in switch regions of the subunits [83, 98].

The high homoacidic and low homobasic content of N-terminal extracellular domains in A-GPCRs (Table 6) indicate a larger ability to contact and dock basic peptide segments, which is in line with highly homobasic constitution of agonists (Tables 1 and S2). This could also help contacts with other ligand-type peptides (Tables 1 and S3–S5).

The transmembrane domains of A-GPCRs are a good illustration for low homoionicity of membrane-crossing segments in proteins. The few transmembrane homoionic tracts represent less than 10 % of domain sequences, and have less than 10 % ionic density (Table 6).

Discussion

Homoionic tracts are not limited to particular ranges of sequence size, and in fact are present even in very short peptides (Tables S2, S4, S5), especially in agonists (Table S2). The presence and abundance of homoionic motifs should be important for contact, docking and stability of association with partners. Homoionicity could be viewed as a basal modularity in polypeptide design (Table 1). In random association, charged homoionic modules could prefer similar counter-charged segments. The frequent stable association of ribosomal proteins and rRNAs obviously depends on use of long basic zippers in the nucleolar rRNA-rich matrix [41, 74]. Stability of heteropentamers formed from GPCR dimers and G-protein heterotrimers [10] should mainly depend on the critical association, in the Gα-rich ER/Golgi [5, 54], of G-protein heterotrimers and the H8 helix of the receptors [95]. This helix essentially is a homobasic zipper [62]. Homoionic modularity could also be a protein construction paradigm, permitting evolutionary modification of complexes with partners by adding or removing same-charge modules. This could apply in particular to highly homobasic sequences of ligand-type polypeptides (Tables 1, S2–S5).

The homoionic modularity points to a number of issues that invite experiments with mutation of biotic sequences and combinatorial sequence design. Examination of the known switching motifs (in e.g. A-GPCRs [8, 87], G-proteins [82, 98], transcription factors [36, 68] and enzymes [71]) in relation to the constituent and interacting homoionic motifs could be of considerable interest, and similar applies to various functionally characterized protein domains, including those shown in Tables 5 and 6.

The lower degree of homoionic segregation for acidic residues relative to basic, apparent in all compared sets, could represent a large-scale evolved pattern related to interactivity of acidic sidechains. Acidic residue-rich peptide segments have the potential to form tight complexes with partner sequences [8, 53]. Even non-contiguous groupings of acidic residues could be powerful interactive motifs [in annexin and Death domains (Table 6)], or adhesive acidic motifs [65]. Proteins rich in acidic zippers, including cadherins, are among the most common protein-linking factors, and motifs with strong anionic tracts regulate traffic and intracellular concentration of divalent cations (e.g. of Ca^{2+} [31]) as well as the protein traffic. Massive multi-acidic zippers are found in a number of transcription factors, and disbanding of their complexes with inhibitors could require proteolysis, as with NFκB/IκB pair [49, 60]. Across large numbers of polypeptide species, an optimized relative homoacidity could be lower than the corresponding homobasicity.

Segregation of basic sidechains, on the other hand, is critical for protein-polynucleotide association (e.g. [34, 58, 86]), as well as for intracellular protein targeting, traffic [23, 70, 77] and organization [92]. Based on ionic density distribution in >9-homoionic zippers, a number of *pset* proteins can have strong (and even blocking) association with the respective partners. Some of these (e.g. protamines) are well known as blockers.

Many agonist peptides, including most chemokine peptides (Table S2), have homobasic tracts with large numbers of ionic sidechains. In the absence of signal transduction, the long homobasic tracts have the potential to make the binding of agonist peptides essentially irreversible, as is documented, e.g. for endothelin at the ET-1 [38] and ET-2 receptor [56], pancreatic polypeptide at the Y4 receptor [61], and PYY(3-36) at the Y2 receptor [25]. Removal of these agonists may require proteolysis. For chemokines, in vitro ligand detachment also indicates a slow separation of agonist peptides [81]. The strong agonist attachment based on homoionic zippers could relate to efficacy of transduction, peptide clearance and nuclear import.

Conclusions

Ionic segregation is implicitly accepted for polynucleotide-interacting proteins, especially protamines and histones, but otherwise is not commonly perceived as potentially general protein feature with low dependence upon type and size. The present study confirms that homoionic sequestration is strongly implemented throughout protein and peptide classes. A sequence differentiation of this type is of interest in examination of protein reactivity, and in evolutionary considerations. This could also be of use in construction of peptidic agonists, antagonists and antibodies.

Acknowledgments This work was partly supported by the U.S. National Institutes of Health grants R01-HD13703 and R01 HD-20074, and also in part by grants from Shriners Hospital for Children, #8570 and #86400 and a Department of Veteran Affairs Grant, 1I01BX000263 (to A.B.).

Conflict of interest The authors declare no conflicts of interest related to subjects of this study.

References

1. Allan J, Harborne N, Rau DC, Gould H (1982) Participation of core histone "tails" in the stabilization of the chromatin solenoid. J Cell Biol 93:285–297
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
3. Annilo T, Karis A, Hoth S, Rikk T, Kruppa J, Metspalu A (1998) Nuclear import and nucleolar accumulation of the human ribosomal protein S7 depends on both a minimal nuclear localization sequence and an adjacent basic region. Biochem Biophys Res Commun 249:759–766
4. Anraku K, Inoue T, Sugimoto K, Kudo K, Okamoto Y, Morii T, Mori Y, Otsuka M (2011) Design and synthesis of biotinylated inositol 1,3,4,5-tetrakisphosphate targeting Grp1 pleckstrin homology domain. Bioorg Med Chem 19:6833–6841
5. Ausiello DA, Stow JL, Cantiello HF, de Almeida JB, Benos DJ (1992) Purified epithelial Na+ channel complex contains the pertussis toxin-sensitive G alpha i-3 protein. J Biol Chem 267: 4759–4765
6. Baker AM, Fu Q, Hayward W, Victoria S, Pedroso IM, Lindsay SM, Fletcher TM (2011) The telomere binding protein TRF2 induces chromatin compaction. PLoS ONE 6:e19124
7. Balhorn R (2007) The protamine family of sperm nuclear proteins. Genome Biol 8:227
8. Ballesteros JA, Shi L, Javitch JA (2001) Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. Mol Pharmacol 60:1–19
9. Ban C, Ramakrishnan B, Ling KY, Kung C, Sundaralingam M (1994) Structure of the recombinant *Paramecium tetraurelia* calmodulin at 1.68 A resolution. Acta Crystallogr D Biol Crystallogr 50:50–63
10. Baneres JL, Parello J (2003) Structure-based analysis of GPCR function: evidence for a novel pentameric assembly between the dimeric leukotriene B4 receptor BLT1 and the G-protein. J Mol Biol 329:815–829
11. Banerjee A, MacKinnon R (2008) Inferred motions of the S3a helix during voltage-dependent K + channel gating. J Mol Biol 381:569–580
12. Baron V, Alengrin F, Van Obberghen E (1998) Dynamin associates with Src-Homology Collagen (Shc) and becomes tyrosine phosphorylated in response to insulin. Endocrinology 139: 3034–3037
13. Barret C, Roy C, Montcourrier P, Mangeat P, Niggli V (2000) Mutagenesis of the phosphatidylinositol 4,5-bisphosphate (PIP(2)) binding site in the NH(2)-terminal domain of ezrin correlates with its altered cellular distribution. J Cell Biol 151: 1067–1080
14. Biddlecome GH, Berstein G, Ross EM (1996) Regulation of phospholipase C-beta1 by Gq and m1 muscarinic cholinergic receptor. Steady-state balance of receptor-mediated activation and GTPase-activating protein-promoted deactivation. J Biol Chem 271:7999–8007
15. Birbach A, Bailey ST, Ghosh S, Schmid JA (2004) Cytosolic, nuclear and nucleolar localization signals determine subcellular distribution and activity of the NF-kappaB inducing kinase NIK. J Cell Sci 117:3615–3624
16. Brennan CM, Gallouzi IE, Steitz JA (2000) Protein ligands to HuR modulate its interaction with target mRNAs in vivo. J Cell Biol 151:1–14
17. Bruno ME, West RB, Schneeman TA, Bresnick EH, Kaetzel CS (2004) Upstream stimulatory factor but not c-Myc enhances transcription of the human polymeric immunoglobulin receptor gene. Mol Immunol 40:695–708
18. Bugatti A, Chiodelli P, Rosenbluh J, Loyter A, Rusnati M (2010) BSA conjugates bearing multiple copies of the basic domain of HIV-1 Tat: Prototype for the development of multitarget inhibitors of extracellular Tat. Antiviral Res 87:30–39
19. Choi JH, Murray JW, Wolkoff AW (2012) PDZK1 binding and serine phosphorylation reg2012ulate subcellular trafficking of organic anion transport protein 1a1. Am J Physiol Gastrointest Liver Physiol 300:G384–G393
20. Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13:222–245
21. Cohen SL, Ferre-D'Amare AR, Burley SK, Chait BT (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. Protein Sci 4:1088–1099
22. Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. EMBO Rep 1:411–415
23. Creancier L, Prats H, Zanibellato C, Amalric F, Bugler B (1993) Determination of the functional domains involved in nucleolar targeting of nucleolin. Mol Biol Cell 4:1239–1250
24. Dang CV, Lee WM (1989) Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. J Biol Chem 264:18019–18023
25. Dautzenberg FM, Neysari S (2005) Irreversible binding kinetics of neuropeptide Y ligands to Y2 but not to Y1 and Y5 receptors. Pharmacology 75:21–29
26. Dessauer CW, Tesmer JJ, Sprang SR, Gilman AG (1998) Identification of a Gialpha binding site on type V adenylyl cyclase. J Biol Chem 273:25831–25839
27. Doyle A, Crosby SR, Burton DR, Lilley F, Murphy MF (2010) Actin bundling and polymerisation properties of eukaryotic elongation factor 1 alpha (eEF1A), histone H2A–H2B and lysozyme in vitro. J Struct Biol 176:370–378
28. Elleuche S, Poggeler S (2011) Inteins, valuable genetic elements in molecular biology and biotechnology. Appl Microbiol Biotechnol 87:479–489
29. Fries M, Chauhan HJ, Domingo GJ, Jung HI, Perham RN (2003) Site-directed mutagenesis of a loop at the active site of E1 (alpha2beta2) of the pyruvate dehydrogenase complex. A possible common sequence motif. Eur J Biochem 270: 861–870
30. Gibb EA, Edgell DR (2011) Better late than early: delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. Mol Microbiol 78:35–46
31. Gifford JL, Walsh MP, Vogel HJ (2007) Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. Biochem J 405:199–221
32. Gogarten JP, Hilario E (2006) Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evol Biol 6:94
33. Grabarek Z (2006) Structural basis for diversity of the EF-hand calcium-binding proteins. J Mol Biol 359:509–525

34. GuhaThakurta D, Draper DE (2000) Contributions of basic residues to ribosomal protein L11 recognition of RNA. J Mol Biol 295:569–580
35. Hallett MA, Lo HS, Bender A (2002) Probing the importance and potential roles of the binding of the PH-domain protein Boi1 to acidic phospholipids. BMC Cell Biol 3:16
36. Hanover JA, Love DC, DeAngelis N, O'Kane ME, Lima-Miranda R, Schulz T, Yen YM, Johnson RC, Prinz WA (2007) The High Mobility Group Box Transcription Factor Nhp6Ap enters the nucleus by a calmodulin-dependent, Ran-independent pathway. J Biol Chem 282:33743–33751
37. Herzog H, Darby K, Ball H, Hort Y, Beck-Sickinger A, Shine J (1997) Overlapping gene structure of the human neuropeptide Y receptor subtypes Y1 and Y5 suggests coordinate transcriptional regulation. Genomics 41:315–319
38. Hilal-Dandan R, Villegas S, Gonzalez A, Brunton LL (1997) The quasi-irreversible nature of endothelin binding and G protein-linked signaling in cardiac myocytes. J Pharmacol Exp Ther 281:267–273
39. Hoeffler JP, Meyer TE, Yun Y, Jameson JL, Habener JF (1988) Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental cDNA. Science 242:1430–1433
40. Hong M, Su Y (2011) Structure and dynamics of cationic membrane peptides and proteins: insights from solid-state NMR. Protein Sci 20:641–655
41. Houmani JL, Ruf IK (2009) Clusters of basic amino acids contribute to RNA binding and nucleolar localization of ribosomal protein L22. PLoS ONE 4:e5306
42. Ikeda M, Arai M, Lao DM, Shimizu T (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. In Silico Biol 2:19–33
43. Jung KC, Rhee HS, Park CH, Yang CH (2005) Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: the inhibitory effect of linoleic acid on the DNA-binding step. Biochem Biophys Res Commun 334:269–275
44. Jung KH, Spudich JL (1996) Protonatable residues at the cytoplasmic end of transmembrane helix-2 in the signal transducer HtrI control photochemistry and function of sensory rhodopsin I. Proc Natl Acad Sci USA 93:6557–6561
45. Klasse PJ, Rosenkilde MM, Signoret N, Pelchen-Matthews A, Schwartz TW, Marsh M (1999) CD4-Chemokine receptor hybrids in human immunodeficiency virus type 1 infection. J Virol 73:7453–7466
46. Komeiji Y, Uebayasi M, Yamato I (1994) Molecular dynamics simulations of trp apo- and holorepressors: domain structure and ligand-protein interaction. Proteins 20:248–258
47. Krebs D, Dahmani B, Monnot M, Mauffret O, Troalen F, Fermandjian S (1996) Dissection of the basic subdomain of the c-Jun oncoprotein: a structural analysis of two peptide fragments by CD, Fourier-transform infrared and NMR. Eur J Biochem 235:699–712
48. Lee JC, Wang GX, Schickling O, Peter ME (2005) Fusing DEDD with ubiquitin changes its intracellular localization and apoptotic potential. Apoptosis 10:1483–1495
49. MacKichan ML, Logeat F, Israel A (1996) Phosphorylation of p105 PEST sequence via a redox-insensitive pathway up-regulates processing of p50 NF-kappaB. J Biol Chem 271:6084–6091
50. Malesevic M, Poehlmann A, Hernandez Alvarez B, Diessner A, Trager M, Rahfeld JU, Jahreis G, Liebscher S, Bordusa F, Fischer G, Lucke C (2011) The protein-free IANUS peptide array uncovers interaction sites between *Escherichia coli* parvulin 10 and alkyl hydroperoxide reductase. Biochemistry 49:8626–8635
51. McPherson PS, Czernik AJ, Chilcote TJ, Onofri F, Benfenati F, Greengard P, Schlessinger J, De Camilli P (1994) Interaction of Grb2 via its Src homology 3 domains with synaptic proteins including synapsin I. Proc Natl Acad Sci USA 91:6486–6490
52. Meskauskas A, Dinman JD (2010) A molecular clamp ensures allosteric coordination of peptidyltransfer and ligand binding to the ribosomal A-site. Nucleic Acids Res 38:7800–7813
53. Mitchell JB, Thornton JM, Singh J, Price SL (1992) Towards an understanding of the arginine-aspartate interaction. J Mol Biol 226:251–262
54. Nakagawa H, Umadome H, Miyazaki S, Tanaka K, Nishimura K, Komori M, Matsuo S (2011) ER-resident Gi2 protein controls sar1 translocation onto the ER during budding of transport vesicles. J Cell Biochem 112:2250–2256
55. Neumann F, Hemmerich P, von Mikecz A, Peter HH, Krawinkel U (1995) Human ribosomal protein L7 inhibits cell-free translation in reticulocyte lysates and affects the expression of nuclear proteins upon stable transfection into Jurkat T-lymphoma cells. Nucleic Acids Res 23:195–202
56. Oksche A, Boese G, Horstmeyer A, Furkert J, Beyermann M, Bienert M, Rosenthal W (2000) Late endosomal/lysosomal targeting and lack of recycling of the ligand-occupied endothelin B receptor. Mol Pharmacol 57:1104–1113
57. Olivares-Reyes JA, Smith RD, Hunyady L, Shah BH, Catt KJ (2001) Agonist-induced signaling, desensitization, and internalization of a phosphorylation-deficient AT1A angiotensin receptor. J Biol Chem 276:37761–37768
58. Palau J, Climent F, Aviles FJ, Morros A, Soliva M (1977) Interactions of histones and histone peptides with DNA Thermal denaturation and solubility studies. Biochim Biophys Acta 476:108–121
59. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289:739–745
60. Palombella VJ, Rando OJ, Goldberg AL, Maniatis T (1994) The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B. Cell 78:773–785
61. Parker MS, Lundell I, Parker SL (2002) Pancreatic polypeptide receptors: affinity, sodium sensitivity and stability of agonist binding. Peptides 23:291–302
62. Parker MS, Park EA, Sallee FR, Parker SL (2011) Two intracellular helices of G-protein coupling receptors could generally support oligomerization and coupling with transducers. Amino Acids 40:261–268
63. Parker MS, Sah R, Balasubramaniam A, Sallee FR, Zerbe O, Parker SL (2011) Non-specific binding and general cross-reactivity of Y receptor agonists are correlated and should importantly depend on their acidic sectors. Peptides 32:258–265
64. Parker MW, Tucker AD, Tsernoglou D, Pattus F (1990) Insights into membrane insertion based on studies of colicins. Trends Biochem Sci 15:126–129
65. Parker SL, Parker MS, Wong YY, Sah R, Balasubramaniam A, Sallee F (2008) Importance of a N-terminal aspartate in the internalization of the neuropeptide Y Y2 receptor. Eur J Pharmacol 594:26–31
66. Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98
67. Pearson WR (2000) Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185–219
68. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG (2004) A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. Cell 116:511–526
69. Pollastri G, Martin AJ, Mooney C, Vullo A (2007) Accurate prediction of protein secondary structure and solvent accessibility

by consensus combiners of sequence and structure information. BMC Bioinform 8:201
70. Romanelli MG, Tato L, Lorenzi P, Morandi C (2003) Nuclear localization domains in human thyroid transcription factor 2. Biochim Biophys Acta 1643:55–64
71. Romano RA, Kannan N, Kornev AP, Allison CJ, Taylor SS (2009) A chimeric mechanism for polyvalent trans-phosphorylation of PKA by PDK1. Protein Sci 18:1486–1497
72. Rosenbaum S, Kreft S, Etich J, Frie C, Stermann J, Grskovic I, Frey B, Mielenz D, Poschl E, Gaipl U, Paulsson M, Brachvogel B (2011) Identification of novel binding partners (annexins) for the cell death signal phosphatidylserine and definition of their recognition motif. J Biol Chem 286:5708–5716
73. Rosenkilde MM, Schwartz TW (2004) The chemokine system—a major regulator of angiogenesis in health and disease. Apmis 112:481–495
74. Rosner M, Fuchs C, Dolznig H, Hengstschlager M (2011) Different cytoplasmic/nuclear distribution of S6 protein phosphorylated at S240/244 and S235/236. Amino Acids 40:595–600
75. Rusnati M, Tanghetti E, Urbinati C, Tulipano G, Marchesini S, Ziche M, Presta M (1999) Interaction of fibroblast growth factor-2 (FGF-2) with free gangliosides: biochemical characterization and biological consequences in endothelial cell cultures. Mol Biol Cell 10:313–327
76. Saudek V, Pasley HS, Gibson T, Gausepohl H, Frank R, Pastore A (1991) Solution structure of the basic region from the transcriptional activator GCN4. Biochemistry 30:1310–1317
77. Shibata Y, Muramatsu T, Hirai M, Inui T, Kimura T, Saito H, McCormick LM, Bu G, Kadomatsu K (2002) Nuclear targeting by the growth factor midkine. Mol Cell Biol 22:6788–6796
78. Singh RN, Saldanha RJ, D'Souza LM, Lambowitz AM (2002) Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. J Mol Biol 318:287–303
79. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow et al. J (eds) Proceedings of the sixth international conference on intelligent systems for molecular biology, vol 1. AAAI Press, Menlo Park, pp 175–182
80. Sperling R, Koster AJ, Melamed-Bessudo C, Rubinstein A, Angenitzki M, Berkovitch-Yellin Z, Sperling J (1997) Three-dimensional image reconstruction of large nuclear RNP (lnRNP) particles by automated electron tomography. J Mol Biol 267: 570–583
81. Springael JY, Le Minh PN, Urizar E, Costagliola S, Vassart G, Parmentier M (2006) Allosteric modulation of binding properties between units of chemokine receptor homo- and hetero-oligomers. Mol Pharmacol 69:1652–1661
82. Sunahara RK, Dessauer CW, Whisnant RE, Kleuss C, Gilman AG (1997) Interaction of Gsalpha with the cytosolic domains of mammalian adenylyl cyclase. J Biol Chem 272:22265–22271
83. Sunahara RK, Tesmer JJ, Gilman AG, Sprang SR (1997) Crystal structure of the adenylyl cyclase activator Gsalpha. Science 278:1943–1947
84. Thiele A, Krentzlin K, Erdmann F, Rauh D, Hause G, Zerweck J, Kilka S, Posel S, Fischer G, Schutkowski M, Weiwad M (2011) Parvulin 17 promotes microtubule assembly by its peptidyl-prolyl cis/trans isomerase activity. J Mol Biol 411:896–909
85. Trub T, Frantz JD, Miyazaki M, Band H, Shoelson SE (1997) The role of a lymphoid-restricted, Grb2-like SH3-SH2-SH3 protein in T cell receptor signaling. J Biol Chem 272:894–902
86. Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci USA 97:11632–11637
87. Vanni S, Neri M, Tavernelli I, Rothlisberger U (2011) A conserved protonation-induced switch can trigger "ionic-lock" formation in adrenergic receptors. J Mol Biol 397:1339–1349
88. Wang X, McLachlan J, Zamore PD, Hall TM (2002) Modular recognition of RNA by a human pumilio-homology domain. Cell 110:501–512
89. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature 347:575–578
90. Weng X, Luecke H, Song IS, Kang DS, Kim SH, Huber R (1993) Crystal structure of human annexin I at 2.5 A resolution. Protein Sci 2:448–458
91. Wiedlocha A, Sorensen V (2004) Signaling, internalization, and intracellular activity of fibroblast growth factor. Curr Top Microbiol Immunol 286:45–79
92. Wolf A, Krause-Gruszczynska M, Birkenmeier O, Ostareck-Lederer A, Huttelmaier S, Hatzfeld M (2012) Plakophilin 1 stimulates translation by promoting eIF4A1 activity. J Cell Biol 188:463–471
93. Xu D, Popov N, Hou M, Wang Q, Bjorkholm M, Gruber A, Menkel AR, Henriksson M (2001) Switch from Myc/Max to Mad1/Max binding and decrease in histone acetylation at the telomerase reverse transcriptase promoter during differentiation of HL60 cells. Proc Natl Acad Sci U S A 98:3826–3831
94. Yamamoto Y, Sono D, Sokawa Y (2000) Effects of specific mutations in active site motifs of 2′,5′-oligoadenylate synthetase on enzymatic activity. J Interferon Cytokine Res 20:337–344
95. Yasuda D, Okuno T, Yokomizo T, Hori T, Hirota N, Hashidate T, Miyano M, Shimizu T, Nakamura M (2009) Helix 8 of leukotriene B4 type-2 receptor is required for the folding to pass the quality control in the endoplasmic reticulum. Faseb J 23:1470–1481
96. Zhou N, Zhang X, Fan X, Argyris E, Fang J, Acheampong E, DuBois GC, Pomerantz RJ (2003) The N-terminal domain of APJ, a CNS-based coreceptor for HIV-1, is essential for its receptor function and coreceptor activity. Virology 317:84–94
97. Zou P, Pinotsis N, Lange S, Song YH, Popov A, Mavridis I, Mayans OM, Gautel M, Wilmanns M (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature 439:229–233
98. Zurita AR, Birnbaumer L (2008) The same mutation in Gsalpha and transducin alpha reveals behavioral differences between these highly homologous G protein alpha-subunits. Proc Natl Acad Sci USA 105:2363–2368