Scaling and self-organized criticality in proteins: Lysozyme *c*

J. C. Phillips

Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA
(Received 21 May 2009; revised manuscript received 29 July 2009; published 20 November 2009)

Proteins appear to be the most dramatic natural example of self-organized criticality (SOC), a concept that explains many otherwise apparently unlikely phenomena. Protein functionality is often dominated by long-range hydro(phobic/philic) interactions, which both drive protein compaction and mediate protein-protein interactions. In contrast to previous reductionist short-range hydrophobicity scales, the holistic Moret-Zebende hydrophobicity scale [Phys. Rev. E **75**, 011920 (2007)] represents a hydroanalytic tool that bioinformatically quantifies SOC in a way fully compatible with evolution. Hydroprofiling identifies chemical trends in the activities and substrate binding abilities of model enzymes and antibiotic animal lysozymes *c*, as well as defensins, which have been the subject of tens of thousands of experimental studies. The analysis is simple and easily performed and immediately yields insights not obtainable by traditional methods based on short-range real-space interactions, as described either by classical force fields used in molecular-dynamics simulations, or hydrophobicity scales based on transference energies from water to organic solvents or solvent-accessible areas.

I. INTRODUCTION

Many physical systems exhibit power-law distributions over limited ranges (hence the popularity of log-log graph paper), and power-law distributions are the characteristic feature of the modern theory of phase transitions near a critical point. Self-organized criticality (SOC) is a methodology that attempts to explain why so many complex systems exhibit power-law distributions and appear to be "accidentally" located near critical points. It is argued that the critical points are dynamical fixed points ("tipping points") toward which the system evolves without tuning external parameters [1]. The critical points are extrema in some property (or properties) with respect to which the system has been optimized, especially with respect to long-range highly cooperative interactions, such as conformational changes.

Given the wide-spread occurrence of power laws, SOC has great intuitive appeal: it has achieved an enduring popularity among theorists (3000+ papers discussing SOC, 25+ current books from a single publisher), notably in modeling the critical stability of sand piles against avalanches (700 papers), but its concrete applications have been limited largely to seismic phenomena (200+ papers). More generally, many synthetic materials have been optimized by trial and error, so it is not surprising to find them near critical points or points where further refinement encounters rapidly diminishing returns. However, there is one family of materials, which has been optimized far beyond all others, and this of course is proteins. The problem has been to find a "handle," which could be used to quantify SOC most simply in proteins.

Apparently, such a handle has been discovered [2] based on amino-acid hydrophobicity, as quantified by statistical studies of trends in solvent-accessible surface areas (SASA), as determined from classical Voronoi partitioning. The unexpected feature of this discovery is that these areas decrease with fragment length in protein fragments containing $(2N+1)$ amino acids, according to power laws in $(2N+1)$

$(1 \le N \le 17)$ with different (centered amino-acid specific) exponents. These exponents derived from bioinformatic scans of 5526 high-resolution protein fragments show that each *aa* induces long-range changes in local backbone curvatures [which are smaller(larger) for hydrophilic(phobic) residues, which are, respectively, exposed or buried in globular proteins]. In general, the fluctuations associated with extrema associated with SOC are expected to involve long wavelengths, such as are involved in large-scale conformational changes in microscopic protein phase transitions.

This discovery [2] immediately suggests SOC, but at the same time it has been regarded skeptically by the biophysics community, the reason being that historically there have been many different definitions of hydrophobicity (although none of the earlier ones involved exponents). Given this skepticism, it seemed natural to put the Moret-Zebende exponent-based hydrophobicity scale Ψ to an independent test, by using it to construct hydrophobic profiles for specific proteins [3]. The proteins chosen were repeat proteins, which consist entirely of α helices connected by short turns, with large SASA. By itself, this is a considerable structural simplification, but repeat proteins are even more special. The amino-acid sequences in successive helices show many sites, where the amino acids are conserved, and the helices themselves have similar lengths. This makes it easy to compare the hydrophobic profiles of successive helices and search for similarities. In this way, many such similarities were found, which could be related to structural properties, including interactions of the repeat proteins with other proteins, for which the former often function as heterotrimeric scaffolds for interprotein interactions or packaging for nucleoporin transport. These similarities can be compared for different hydrophobicity scales: as expected from SOC, the similarities and their relation to structural properties were always maximized using the MZ SOC Ψ scale. This "kills two birds with one stone:" it verifies the accuracy of the MZ SOC Ψ scale (which may be virtually exact), and it also demonstrates that proteins are by far the most elegant and interesting example of SOC.

II. LYSOZYMES

The next question is whether or not repeat proteins (<1 repeat %) archea to (>5 repeat %) metazoa [3] are the only proteins where hydrophobicity is the dominant functional factor. Here, we study lysozyme c and show that it too can be analyzed rewardingly with the MZ SOC Ψ scale. Lysozymes are a ubiquitous protein family [4], which contains hen egg-white (HEW) lysozyme, up to 1985 the most studied protein. Lysozymes function both as enzymes and as antibiotics. The enzyme function is much studied and is regarded as archetypical. It was originally (1965) supposed to be mediated by short-range ionic interactions; it has only recently (2001) been identified as also mediated by both short-range covalent and H-bond interactions [5]. Antimicrobial (AMB or lytic) mechanisms are much more complex and are still the subject of many studies [6]. Although HEW lysozyme (129 residues) is much larger than defensins ($\sim$30 residues), it shares essential structural and AMB functions. Using the Moret-Zebende hydrophobicity scale Ψ based on the long-range power-law evolution of SASA with increasing segmental length, one can compare species trends in both enzyme and AMB activity of defensins, lysozymes, and related proteins. The results for protein functionality obtained with the holistic long-range MZ scale are consistently superior to those obtained from short-range reductionist Ψ scales based on transference energies of isolated amino acids from water to organic solvents (KD scale) or unevolved SASA (BF scale) [7,8]; the differences are especially pronounced in functionally critical regions [9,10]. The insights obtained can be used to engineer new proteins with potentially desirable AMB properties [11]; such insights are expected on fundamental grounds, as the first hydration monolayer exhibits distinctive properties [12].

III. SYSTEMATICS OF WILD LYSOZYMES

Much of the source material used here comes from two review articles [13,14]. Most discussions of enzyme functionality begin with the traditional lock and key mechanism, which emphasizes short-range interactions (sometimes based simply on charge transfer), a picture that has been refined by successive mutagenic studies of enzyme cores, often complexed with a simple substrate. Thus, the basic structure of lysozyme c consists of two lobes or domains (right, α helices; left, β strands) surrounding an active site cleft (Fig. 1), which binds six sugar rings (A-F). Modeling led to the insight that the nearby conserved basic acid pair (left) Asp^{52} and (right) Glu^{35} could exchange charge to stabilize the transition state (or unstable intermediates) in the reaction mechanism [5]. This primary model has since been supported in many ways; here it is revisited to determine the secondary long-range factors that determine quantitative species trends in wild lysozymes [13]; lists the amino-acid sequences of 75 animal and insect c lysozymes, from HEW to human; detailed study of these, primarily supported by the known structure and the MZ hydrophobicity Ψ scale, and secondarily by other scales (such as less accurate and contextually limited helical propensity scales [15]) enables us to recognize long-range interactions that are important for protein
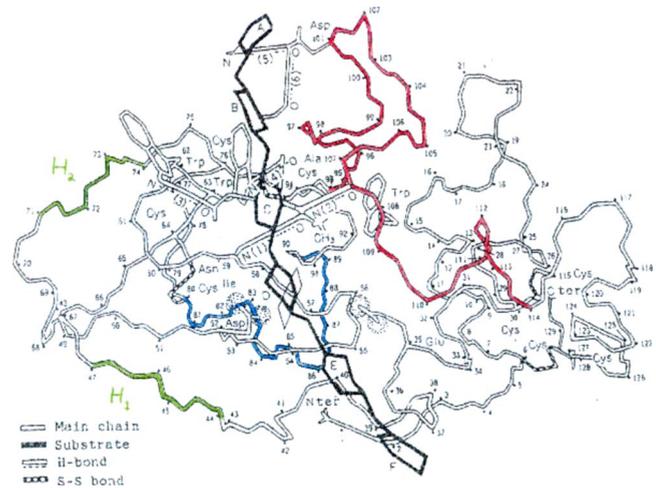
FIG. 1. (Color online) A sketch of lysozyme c, adapted from [14]. Six sugar rings (A-F) are indicated in the cleft between two lobes (left, α helical and rigid; right, β strands and flexible). The basic mechanism for enzyme activity is still charge exchange between conserved Glu 35 and Asp 52, but most of the other features (such as the emphasis placed on Trp) inherited from [13,14] no longer appear to be significant [15]. The new points are explicit identification of the factors responsible for the species-dependent chemical trends shown in Table I. These include the hydrophobically skewed scissors 80–94 (blue online), the cationic-rich flexible right lobe segment between C94 and C115 (red online, and the two hydrophobically adjusted left lobe ears $H_1 = 44-47$ and $H_2 = 71-74$ (green online). Most of the trends in animal lytic activity are associated with amphipathic changes in the net charge of these two ears and of the cationic-rich segment.

engineering. Equally important are the short-range interactions studied by mutations [14]; in practice, one would engineer a combination dictated by these and other factors. Because of competing interactions, hydrophobicity profiles typically exhibit waves, with lengths between extrema of order 5–6 residues. In analyzing such profiles, it is natural for one to focus on extrema, a method that appears to work well.

The lysozyme species trends studied here concern enzymatic binding ability against three sugar rings [N-acetylglucosamine, (NAG)$_3$], as well as AMB or lytic activity against gram-positive bacteria. These are shown in Table I reproduced here for the reader's convenience from [14], for seven examples. These examples fall naturally into two groups (birds and placental mammals), and when these are arranged in order of lytic activity, the primary evolutionary groups are simply separated. This is not the case for enzymatic binding ability (necessary for metabolic activity against glycol chitin), and it is clear that evolution has aimed mainly at strengthening lytic activity [6], which is more complex than metabolic activity. The very large binding ability of HEW lysozyme against multiple sugar rings is striking: it appears that the HEW sequence is nearly perfectly adapted to this task, while the turkey lysozyme sequence (presumably developed 30 million years later [16]), only slightly changed from HEW, is much weaker.

Of the 130 lysozyme c residues, 23 are invariant among animals, including 8 Cys residues that form 4 disulfide bonds

TABLE I. The entries for lytic activity (against gram-positive bacteria) and activity against glycol chitin (a soluble linear homopolysaccharide) are normalized to 100% for HEW [from (14)]. The sugar-binding ability is represented by the association constant K_A against $(NAG)_3$. The three- and four-site average extremal (Cys-based) hydrophobicities are calculated with the "exact" SOC MZ scale (columns 5 and 6) and the reductionist KD scale (columns 7 and 8) for two subsets (M and M*) of pheas. and turkey mutations relative to HEW (see text).

Lysozyme	Lytic act.	Glycol chitin act.	Sugar bind.	$\langle\langle\Psi3\rangle_M\rangle$	$\langle\langle\Psi4\rangle_{M^*}\rangle$	$\langle\langle\Psi3\rangle_M\rangle$	$\langle\langle\Psi4\rangle_{M^*}\rangle$
HEW	100	100	71400	0.143	0.143	0.148	0.133
Pheas.	123	82	55600	0.168	0.154	0.169	0.138
Turkey	176	80	8300	0.159		0.161	
Rabbit	204	99	17500				
Pig I	245	45	8300				
Rat 1b	255	99	9200				
Human	396	110	10000				

[13], as well as the basic acid pair Asp[52] and Glu[35] [13], identified a region around invariant Trp residues 108 and 111, as well as 28, as a "hydrophobic region A," and a long loop sequence 50–76 as a "hydrophilic domain B." In general, we will average an amino-acid hydrophobicity Ψ over a $(2N+1)$ window to obtain a contextual average $\langle\Psi2N+1\rangle$. On the MZ scale, Trp is not strongly hydrophobic (it is merely a typically hydrophobic "core" residue), and the three-residue primitive contextual averages [3] over X=Trp 28, 108, and 11 of MZ $\langle\Psi3\rangle(X)$ are 0.167 (HEW) and 0.165 (human), in other words, little changed. While important for stability, Trp is not critical to lysozyme functionality [17]. Domain B (the long loop sequence 50–76) is indeed hydrophilic in HEW ($\langle\Psi27\rangle=0.147$), but in human lysozyme, $\langle\Psi27\rangle=0.153$ [quite close to hydroneutral (0.155)], so this loop is possibly a factor in the large HEW sugar-binding ability, which has weakened in human lysozyme.

IV. WILD HEW, TURKEY, AND RN (JAPANESE) PHEASANT LYSOZYMES

We can test ideas, such as the importance of the hydrophilicity of domain B, by combining the MZ scale (with its long-range SOC accuracy) with the detailed changes in properties of these three birds, as listed in Table I. As expected, there are few sequence changes between them, and only two single mutations in domain B. Overall, most of the mutations are singles, many of which are innocuous [for example, HEW R73→(TUR and PHE)K73 does not alter charge, hydrophobicity, or molar volume significantly]. There is virtually no change in $\langle\Psi27\rangle$ for domain B between HEW, TUR, and PHE, which means that after all this loop is not an important factor in weakening the large HEW sugar-binding ability [which decreased by approximately a factor of 9 between HEW and TUR (Table I)].

There are three significant HEW→TUR contextual hydrophobicity changes: RH[15]G→RL[15]G, which changes $\langle\Psi3\rangle(15)$ from 0.129(HEW) to 0.144(TUR); TQ[41]A →TH[41]A, which changes $\langle\Psi3\rangle(41)$ from 0.132(HEW) to 0.148(TUR), and VQ[121]A→VH[121]A, which changes $\langle\Psi3\rangle(121)$ from 0.167(HEW) to 0.182(TUR). Individually,

these changes are all small (because only one site is mutated, and $\langle\Psi3\rangle$ is averaged over three sites), but they all have the same sign and are in fact nearly equal (the TUR sites are all more hydrophobic). These three sites lie well outside the sugar-binding cleft and the hydrophobic core on the protein surface. The nearly perfect compaction of HEW (400 million years old [4]) by hydrophobic forces suggests that the water monolayer of HEW is an exceptionally well-adapted (stress-free) glassy network, similar to the nearly ideal networks observed in the reversibility windows of network glasses [18]. Another interpretation of this near equality of hydrophobicity changes is that the interfacial water-protein interfacial tension is almost constant around the entire protein surface (as it has to be if the protein sequence is evolutionarily adapted to optimize these interactions, and as Gibbs would have hoped for his droplet nucleation model). These three hydrophobic mutations in TUR cooperatively disrupt the HEW water monolayer network and thereby stiffen the TUR backbone, so that its ability to bind multiple sugar rings is much weaker than HEW's. At the same time, the lytic activity of TUR lysozyme is enhanced (Table I), probably because this depends on large-scale lysozyme cooperativity.

The next case is HEW→PHE. Here there are seven small but significant mutations, and just as with TUR, they occur outside the hydrophobic core on the protein surface. All seven changes are small and again of the same sign, much as for HEW→TUR, implying strong support for the interfacial water-protein surface-tension mechanism discussed above. Thus, one would expect to see differences in properties of HEW and PHE to be twice as large as for TUR and HEW; but according to Table I, they are only about 1/3 as large. What has happened? The answer is that these seven mutations have separated into two subsets, with four larger mutational $\langle\Psi3\rangle$ changes distant from Cys disulfide bridges and three smaller mutational $\langle\Psi3\rangle$ changes adjacent to bridges (denoted by *). The four distant HEW→TUR larger mutational $\langle\Psi3\rangle$ changes are 15(129→152), 41(132→149), 121(167→196), and 122(145→175), while the three adjacent smaller mutational $\langle\Psi3\rangle$ changes are 77*(194→207), 114*(90→100), and 115*(146→156).

Disulfide bridges cause otherwise singly connected protein chains to be multiply connected. Modified hydrophobic

TABLE II. Amino-acid hydrophobicities [KD (7), BF (8), MZ (2)]; the first two scales have been linearly rescaled so that their averages and ranges match those of the MZ scale. Thus, the two most hydrophobic amino acids are Ile and Val (KD scale), Cys and Ile (BF scale), and Cys and Val (MZ scale). The MZ scale, based on SOC, consistently reveals more physiologically significant details. The correlation coefficients of the KD (7) and BF (8) scales with the MZ (2) scale are 0.848 and 0.935, respectively. The BF scale is the most widely used hydrophobicity scale at present.

A	0.200	C	0.214	D	0.096	E	0.096	F	0.220
	0.156		0.230		0.103		0.103		0.217
	0.157		0.246		0.087		0.094		0.218
G	0.157	H	0.101	I	0.253	K	0.088	L	0.240
	0.164		0.173		0.217		0.059		0.204
	0.156		0.152		0.222		0.069		0.197
M	0.202	N	0.096	P	0.133	Q	0.096	R	0.076
	0.204		0.107		0.112		0.103		0.112
	0.221		0.113		0.121		0.105		0.078
S	0.149	T	0.157	V	0.248	W	0.147	Y	0.139
	0.121		0.138		0.208		0.203		0.164
	0.100		0.135		0.238		0.174		0.222

interactions near disulfide bridges can strengthen these bridges and increase the ability of PHE relative to TUR to bind sugar rings, thus, bringing it closer to HEW in properties. In fact, one can define a mutational configuration coordinate M^{**}

$$\langle\langle\Psi3\rangle_M^{**}\rangle = \langle\langle\Psi3\rangle_M\rangle - 2\langle\langle\Psi3\rangle_{M^*}\rangle, \quad (1)$$

with $\langle\Psi3\rangle_M$ averaged over mutations distant from disulfide bridges, and $\langle\Psi3\rangle_{M^*}$ averaged over mutations adjacent to bridges. The factor of 2 in Eq. (1) is what one would expect if the hydrogen bonds involved in water monolayer-protein interfaces are stabilized quantum mechanically and are out of phase at branching disulfide bridges, or if internal stresses are balanced at the multiply connected bridging sites. Sugar binding correlates well with $\langle\langle\Psi3\rangle_M^{**}\rangle$ (Table I).

The discussion leading to Eq. (1) is quite abstract, and it may seem unjustified to many readers; indeed, like much abstract mathematics, it requires much more time to understand than to read. The key points are the nearly equal hydrophobic steps between HEW and PHE that occur for different wild sequences at seven spatially distant superficially unrelated sites on an essentially common lysozyme surface: this is the hydrophobic analog of the classical concept of protein water interfacial surface tension. (This concept assumes that the tension is nearly constant over the interface, which is necessary if this interaction makes the dominant contribution to stabilizing the main features of the surface geometry.) As in Gibbs nucleation models of first-order phase transitions, surface or interfacial tension is expected to be an essential factor determining functional properties of proteins regarded as self-organized networks near critical points. Such equality is of course very unlikely *a priori*, but one can go further. The MZ scale itself is based on exponents from power-law fits to the long-range length dependence of solvent-accessible surface areas (2), so that its underlying justification is SOC: this is the unifying holistic mechanism that leads to nearly equal hydrophobic steps for spatially

distant sites on an essentially common lysozyme surface.

To test this idea, one can repeat the profiling using one of the many reductionist hydrophobicity Ψ scales based on transference energies of individual amino acids from water to an organic solvent [7,8]. Experience has shown that reductionist Ψ scales generally give qualitative trends that are similar to those obtained with the holistic MZ Ψ scale, but they lack the details that provide convincing models of protein functionality. So it is here, the seven nearly equal steps found with MZ become widely unequal, and sometimes even reverse sign, with the KD scale [7]; it appears that the overall signal/noise ratio has dropped by at least a factor of 4 from the unifying holistic MZ scale to the fragmented reductionist KD scale and experience with other reductionist examples suggests that their results would be equally noisy [8].

For completeness, the final KD results are included in Table I, and it might appear that they are quite similar to those with the MZ scale. However, this qualitative similarity is deceptive: it occurs because these abbreviated results are based on multiple averages. The key point of equal steps for distant surface sites for the MZ scale, but not for the KD scale, is not shown here, but it can easily be checked by the reader using hydrophobicity tables given in Table II. One cannot infer Eq. (1) from these multiply averaged results, but it does appear to be quite natural when one studies the nearly equal individual contributions to the two subsets $\langle\langle\Psi3\rangle_M\rangle$ and $\langle\langle\Psi3\rangle_{M^*}\rangle$ (but only with the MZ scale, not with the KD or BF scales).

V. DISULFIDE BONDS IN HEW AND HUMAN LYSOZYME

Compared to HEW, human lysozyme has 72 conserved sites. The 58 mutations include many innocuous ones, but even so the number of mutations is too large to be treated merely as perturbations, as were those of turkey and pheasant. Instead, we focus on comparing the hydro(phobic, philic) extrema. On various reductionist hydrophobicity Ψ scales (Table II), the most hydrophobic amino acids are ei-

TABLE III. Hydroanalysis of contextual disulfide (Cys-Cys) bonds in human lysozyme using the holistic MZ and reductionist KD scales. The first four C's are bonded to the last four.

Peak	$\langle \Psi 3 \rangle$ MZ	MZ C strength	$\langle \Psi 3 \rangle$ KD	KD C strength
MC30L	221.3	Strong	218.8	Strong
YW64C	214	Strong	167.0	Weak
VA94C	213.7	Strong	220.8	Strong
CG129V	213.3	Strong	206.4	Strong
CE7L	179	Weak	183.4	Weak
CH78L	198.3	Weak	185.4	Weak
LS80C	181	Weak	201.0	Strong
CQ117N	154.7	Weak	135.4	Weak

ther Ile, Trp, or Cys [4,7,8]. Because Cys is the most hydrophobic amino acid in the MZ SOC Ψ scale, with $\Psi(C) = 0.246$, most of the MZ hydrophobic extrema are associated with disulfide C-C bonds. It is likely that such hydrophobic extrema are responsible for the special properties that have attracted the greatest interest to Cys-rich proteins. In the last 25 years, there have been Y papers on X-rich proteins, with $(X, Y) = (\text{Cys}, 4900)$ [most hydrophobic]; (Leu, 3700) [hydrophobic]; (Ala, 760) [hydroneutral]; and (Lys or Glu, 500) [hydrophilic]. However, we do not list $\langle \Psi 3(C) \rangle$ for each C but rather the contextual hydrophobic extremum of $\langle \Psi 3(Y) \rangle$, where C may be either X, Y, or Z of an XYZ sequence; in other words, C can be at the center of the three-residue extremal sequence or a nearest neighbor. [Notice that these $\langle \Psi 3(C) \rangle$ extrema will usually have values considerably smaller than $\Psi(C)$ but still well above hydroneutral $\Psi(\text{Ala}) = 0.155$].

The advantage of concentrating on these contextual C hydrophobic extrema is brought out by comparing the eight extremals for HEW and human lysozyme (Table III). In HEW, the extrema show no special properties, but in human lysozyme they separate nicely into two groups: **strong*** ($\langle \Psi 3 \rangle > 0.213$) and weak ($\langle \Psi 3 \rangle < 0.198$) extrema. Is that all? No there is more, and it is spectacular. The two human subgroups pair off exactly, with disulfide bonds formed only between (unlike) **strong*** and weak extrema [(6–**128***), (**30***–115), (**65***–81), and (77–**95***)]. Thus, the four disulfide bonds of human lysozyme are maximally similar (in a set-theoretic {**strong***}/{weak} pairing sense), whereas the four disulfide bonds of HEW exhibit no special properties. This complete separation of the eight C's of human lysozyme into two cross-bonded and remotely ordered subgroups could provide the simplest indication of the evolutionarily optimized cooperative mechanisms responsible for the maximal lytic activity of human lysozyme.

Comparison of these human lysozyme results obtained with the holistic MZ Ψ scale with those obtained with the reductionist KD Ψ scale brings out their significance (Fig. 1). Again with the KD scale, there are two groups of four extrema each, there is a gap between strong ($\langle \Psi 3 \rangle > 0.201$) and weak ($\langle \Psi 3 \rangle < 0.185$) extrema groups, and it has a similar magnitude. However, this time YW[64]C (which was in the strong group with the MZ scale) has switched to the weak

group, while LS[80]C (which was in the weak group with the MZ scale) has switched to the strong group. Thus, the four disulfide bonds consist of two bonds each between like and unlike, or like and like, and with the KD scale there is no qualitative difference between the disulfide bonds of HEW and human lysozyme. One can say that (so far as disulfide cross-linking is concerned), it has taken evolution 400 million years fully to embed SOC in lysozyme c through the strong/weak disulfide bond pairing.

One might suppose that the simple geometrical hydrophobicity Ψ scale [8] based on SASA buried on folding (BF, which is the most popular scale) would give results similar to those of the MZ scale based on SOC because the correlation coefficient of BF with MZ is 0.935 compared to only 0.848 for KD. As expected from the correlation coefficients, the disulfide results of BF are halfway between those of MZ and KD. The different gaps between strong and weak disulfide bond pairs disappear (Fig. 1), and the BF Ψ scale is neither well paired (MZ) nor mispaired (KD); it is simply inconclusive.

Because Cys has the largest hydrophobicity Ψ in the MZ scale, disulfide bonds are ideal markers for hydrophobic lysozyme extrema. The situation is less simple for hydrophilic extrema. When one plots $\langle \Psi 3 \rangle$ (MZ) for HEW and human lysozyme (not shown here), there are long sequences of good agreement, even though the individual amino acids are often different. The two outstanding patches of hydrophilic disagreement are 42–44, where $\langle \Psi 3 \rangle \sim 0.15$ (human, hydroneutral) and 0.10 (HEW, hydrophilic), and 71–74, where $\langle \Psi 3 \rangle \sim 0.17$ (human, hydrophobic) and 0.115 (HEW, hydrophilic). Both patches are on the surface of the left lobe, and both involve hydrophobic stiffening of the human lysozyme surface patches relative to those of HEW lysozyme. This stiffening may contribute to stabilizing cooperative (dimerized) human lysozyme-lysozyme interactions in lytic activity (see below).

VI. WILD HUMAN, RAT, PIG 1, RABBIT, AND MOUSE LYSOZYMES

Next, one can analyze these five placental cases, using the best-studied case (human) as benchmark. Our simplest configuration coordinate is the separation of disulfide bonds into two subgroups, which then leads to four **strong***/weak human pairs. In rabbit, rat 1b and pig 1 lysozyme, C[77]YZ switches from weak to strong, spoiling one pair in rabbit. However, in rat 1b and pig 1, XYC[95], which is bonded to C[77]YZ, also switches from strong to weak, restoring this pair, so that rat 1b still has four strong-weak pairs. Finally, a third switch occurs in pig 1 (C[128]YZ from strong to weak), once again spoiling one pair. Thus, human and rat 1b have four contextual disulfide bonds each between hydrophobically strong and weak CYs, while mammalian rabbit and pig 1 have only three such bonds. These relations are not evolutionary, they are functional. Looking at Table I, we see (as already guessed above) that the numbers of these bonds correlates well with trends in lytic activities of these four species. Because of the growing importance of mouse genomically, note that mouse M (macrophage) and mouse P (small
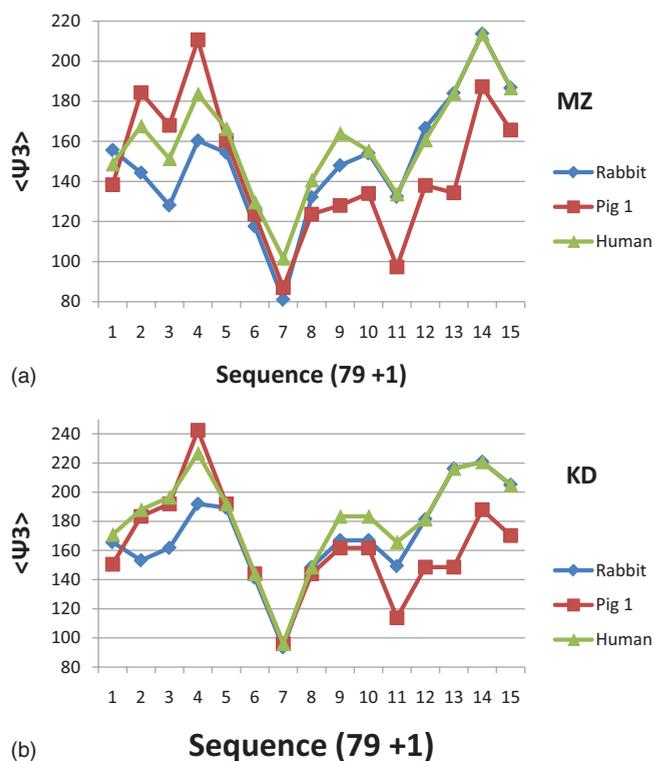
FIG. 2. (Color online) (a) MZ $\langle\Psi3\rangle$ hydropatterns between Cys80 and Cys94 (the interlobe scissors in Fig. 2). (b) KD $\langle\Psi3\rangle$ hydropatterns between Cys80 and Cys94 (the interlobe scissors in Fig. 2).

FIG. 3. (Color online) Contextual hydrophobicities $\langle\Psi3\rangle$ of seven cysteines (at sites indicated at right) involved in four disulfide bonds in human lysozyme, according to three different scales (KD, BF, and MZ, see text and Table II.). The eighth contextually least hydrophobic cysteine (116) lies below the range shown ($\langle\Psi3\rangle(116) \sim 0.145$) and is weakly hydrophilic in all scales. There is a gap between the upper and lower four cysteines for the KD and MZ scales, but not the BF scale. By utilizing this gap, one can identify strong/weak Cys $\langle\Psi3\rangle$ pairings that are maximized in human lysozyme c (see text). The crossing pattern shown here reflects the gains in accuracy as one proceeds from KD to BF to MZ.

intestine) have the same "perfect" lysozyme disulfide pairing as human and rat 1b.

As we would expect, except for this almost hidden disulfide bond correlation, there are long sequences, where differences between human and the other three species hydrophobicities are small. Between human and rat 1b, one notices two interesting differential patches; for 19–21, human $\langle\Psi3\rangle$ is close to 0.15 (hydroneutral), whereas rat 1b is close to 0.20 (hydrophobic, comparable to Cys). Here mouse M is the same as rat 1b, while mouse P is the same as human. Then for 71–73, the cationic charge pattern is different: human 0, HEW and mouse M, +1, and rat and mouse P, +2.

When we compare mammalian rabbit, pig 1, and human lysozymes, we again find long sequences, where differences between human and the other two mammalian species hydrophobicities are small. However, the differences in the long sequence C^{80}-C^{95} are striking. This sequence is an interlobe "scissors," wrapped around the glycol bond "cord" between the D and E sugar rings of lysozyme-$(NAG)_6$ complexes, and it is shown in blue (online) in Fig. 2. Figures 3(a) shows MZ "blue scissors" hydrophobicities, and we see that human and rabbit hydrophobicities are similar; but pig 1 is significantly different, especially on the right lobe (88–94). In the first half of the sequence, pig 1 is more hydrophobic, and in the second half it is more hydrophilic. The water imbalance between the lysozyme left lobe (first half of sequence) and its right lobe (second half) is thus altered for pig 1 compared to human and rabbit. Next, we look at Table I, and we find that the glycol chitin activity of pig 1 is half that of the other species.

Following such a spectacular correlation for the MZ holistic scale, one naturally asks, how successful is the KD result for hydrophobic trends in the long scissors sequence C^{80}-C^{94}? The KD trends are shown in Fig. 3(b). At first glance, the results obtained with the reductionist KD scale look quite similar to those obtained with the holistic MZ scale, but close inspection reveals crucial differences. True, the two scales give similar trends between the three species for the second half (right lobe) of the sequence, but this second half success is canceled by failure with the first half (left lobe). Using the reductionist Ψ scale alone, one would probably not be able to recognize the hydrophobic twisting of the rabbit sequence, which is correlated with its halved glycol chitin activity. Note that this scissors sequence plays a secondary role (compared to the primary role of the conserved basic acid pair Asp^{52} and Glu^{35}), but it is just such subtle *nonconserved* hydrophobic effects that one cannot identify except with the holistic MZ exponent Ψ scale.

VII. LYTIC ACTIVITY

Cationic residues are believed to be the largest factor in determining antimicrobial activity of peptide segments [19,20], which are also amphipathic. The cationic-rich segment of lysozyme lies between C^{95} and C^{115} (see red segment in Fig. 1), and fragments of this segment exhibit lytic activity [20,21]. The net charge (K+R-D-E) in HEW and phea is +3, turkey +4, rabbit +3, (pig 1 and rat 1b) +4, and this increases to +5 in human lysozyme: within the seven-

membered lysozyme family of Table I, the net charge of this segment correlates very well with lytic activity (except for rabbit).

Because the cationic-rich segment of lysozyme lying between C^{95} and C^{115} is located on the right lobe of the cleft in Fig. 1, one is tempted to guess that these two lobes themselves are globally amphipathic. The left lobe of human lysozyme (41–86) has $\langle \Psi 46 \rangle = 0.153$, while the right lobe is more hydrophilic, with $\langle \Psi 84 \rangle = 0.148$, so the guess seems to work. However, proteins are full of surprises. In pig 1, the two lobe hydrophobicities are nearly equal, and when one examines HEW, rabbit and rat 1b lysozymes, one finds a reverse relation, with left lobe values 0.145 ± 0.001 and right lobe values 0.151. [Some might worry that these results are an artifact of the holistic MZ scale, but a similar (albeit smaller) reversal occurs even with the reductionist KD Ψ scale; as usual, the MZ Ψ scale is much more accurate and shows a larger effect.]

It appears that the amphipathic left-right lobe reversal reflects fundamental differences in lysozyme functionality. In HEW, rabbit and rat 1b lysozymes, different amphipathic mechanisms are operative, which switch over to become the global lobe mechanism in human lysozyme, which thus differs substantially from HEW both in net charge and amphipathicity. The largest hydrophobicity differences occur between HEW and human lysozyme in two regions H_1 and H_2: HEW $N^{44}RNT^{47}$ ($\langle \Psi 4 \rangle = 0.110$) is much more hydrophilic than human $N^{44}YNA^{47}$ ($\langle \Psi 4 \rangle = 0.151$), and HEW $G^{71}SRN^{74}$ ($\langle \Psi 4 \rangle = 0.111$) is much more hydrophilic than human $G^{71}AVN^{74}$ ($\langle \Psi 4 \rangle = 0.166$). Moreover, the double HEW R^{45} →human Y^{45} and HEW R^{73}→human V^{73} exchanges reduce the positive charges in both H_1 and H_2 of human lysozyme, thus, enhancing the effectiveness of the dipolar (cationic, hydrophobic) synergistic lytic mechanism for porin (or holin) formation (6). In the other animal lysozymes, the hydrophilic R positive charge is restored in H_2 by human G^{71}→(rabbit, rat 1b and pig 1)R^{71}, thus, producing an intermediate amphipathicity with only one R exchanged region H_1. It is both striking and surprising that evolution did not alter the global lobe amphipathic mechanism gradually (as suggested by Table I) nor abruptly between birds and placental animals (as one might have expected on traditional biological grounds), but instead left most of the animals with only an HEW H_1 local mechanism and provided only humans (not even mice) with the global lobe mechanism with two localized H_1 and H_2 left lobe regions.

VIII. SYNTHETIC POINT MUTATIONS

Mutational experiments have been performed to support and analyze the details of the nearby basic acid pair Asp^{52} and Glu^{35} interactions [5], all with spectacular success. Generally speaking, modifications of either residue are sufficient to destroy both lytic and enzymatic activity. A number of mutational studies have altered lysozyme stabilities (and

sometimes even enhanced them), but it is rare for activities to be enhanced rather than destroyed [14,19]. This is scarcely surprising, as the proteins have evolved to optimize their activities while maintaining merely sufficient stability. Thus, the cores of lysozyme proteins are nearly perfectly conserved, while species trends in wild-type protein properties (previously attributed to unspecified differences in sequences [14] and internal backbone stiffness [21,22]), are caused by long-range hydrophobic interactions of the types identified here with the holistic MZ Ψ scale.

IX. BOVINE α-Lactalbumin and Human Defensins

These cases are discussed elsewhere [23].

X. CONCLUSIONS

The differences between HEW and human lysozymes c (as well as the other five animal lysozymes discussed here) are inaccessible to most theoretical structural probes, not only molecular-dynamics simulations using classical force fields, but also even more sophisticated "soft mode" or principal component dynamical methods designed to identify domains and hinges, as the C_α coordinates of these lysozymes are superposable to 0.65 A [21]. As discussed in [24], there are profound differences between stability and functionality. Conserved sites, including even conserved disulfide and salt bridges, as well as 30–40 % sequence conservation, yielding almost identical backbone folds, primarily assures protein stability within a given family of proteins (lysozymes or defensins), while leaving unexplained species trends in functionality associated with nonconserved sites.

The MZ hydrophobicity Ψ scale includes the effects of self-similarity and self-organized criticality, and this enables it to explain species trends in functionality that are simply inaccessible to most theoretical structural probes, as well as less accurate Ψ scales that describe protein stability and transition states associated with dysfunctional protein unfolding. Scale-free and small-world properties, SOC topological characteristics, have also been recognized in small globular protein studies [21]. However, the successes described here and elsewhere leave an important question unanswered: just why are proteins so close to SOC? One can conjecture that the power-law evolution of solvent-accessible surface areas describes compacted yet still stress-free protein networks softened by water [10,24,25]. This mathematical aspect of evolution is possible because the richness of the amino-acid menu has made it possible for proteins to adapt their (several) functionalities optimally and specifically through water-mediated interactions. Applications of the MZ hydrophobicity Ψ scale to repeat proteins also show its superiority to alternative scales [7,8]. Although the number of papers discussing SOC is ~3000 at this writing, until now there are only a few applications [3] of this powerful concept in biophysics, and many others remain to be made, in this subject that is far more interesting than sand piles.

[1] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).

[2] M. A. Moret and G. F. Zebende, Phys. Rev. E **75**, 011920 (2007).

[3] J. C. Phillips, Proc. Natl. Acad. Sci. U.S.A. **106**, 3107 (2009); **106**, 3113 (2009).

[4] P. Jolles, *Lysozymes: Model Enzymes in Biochemistry and Biology* (Birkhauser Verlag, Basel, 1996).

[5] D. Ringe and G. A. Petsko, Science **320**, 1428 (2008).

[6] T. Park, D. K. Struck, J. F. Deaton, and R. Young, Proc. Natl. Acad. Sci. U.S.A. **103**, 19713 (2006).

[7] J. Kyte and R. F. Doolittle, J. Mol. Biol. **157**, 105 (1982).

[8] G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus, Science **229**, 834 (1985).

[9] P. A. Alexander, Y. He, Y. Chen, J. Orban, and P. N. Bryan, Proc. Natl. Acad. Sci. U.S.A. **104**, 11963 (2007).

[10] A. E. Kister and J. C. Phillips, Proc. Natl. Acad. Sci. U.S.A. **105**, 9233 (2008).

[11] T. Durek, V. Y. Torbeev, and S. B. H. Kent, Proc. Natl. Acad. Sci. U.S.A. **104**, 4846 (2007).

[12] F. Merzel and J. C. Smith, Proc. Natl. Acad. Sci. U.S.A. **99**, 5378 (2002).

[13] E. M. Prager and P. Jolles, *Lysozymes: Model Enzymes in Biochemistry and Biology* (Birkhauser Verlag, Basel, 1996).

[14] T. Imoto, *Lysozymes: Model Enzymes in Biochemistry and Biology* (Birkhauser Verlag, Basel, 1996).

[15] C. N. Pace and J. M. Scholtz, Biophys. J. **75**, 422 (1998).

[16] D. K. Griffin *et al.*, BMC Genomics **9**, 168 (2008).

[17] M. Stuart-Audette, Y. Blouquit, M. Faraggi, C. Sicardi-Roselli, C. Houee_Levin, and P. Jolles, Eur. J. Biochem. **270**, 3565 (2003).

[18] S. Chakravarty, D. G. Georgiev, P. Boolchand, and M. Micoulaut, J. Phys.: Condens. Matter **17**, L1 (2005).

[19] R. E. W. Hancock, Lancet **349**, 418 (1997).

[20] T. S. Ryge and P. R. Hansen, J. Pept. Sci. **11**, 727 (2005).

[21] S. Hayward, A. Kitao, and H. J. C. Berendsen, Proteins **27**, 425 (1997).

[22] N. Hori, G. Chikenji, R. S. Berry, and S. Takada, Proc. Natl. Acad. Sci. U.S.A. **106**, 73 (2008).

[23] J. C. Phillips, e-print arXiv:0808.2312.

[24] J. C. Phillips, e-print arXiv:0802.3641.

[25] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Nat. Phys. **4**, 711 (2008).