

Structural and electrostatic properties of ubiquitination and related pathways

Peter J. Winn¹, Mai Zahran^{1,2,3}, James N. D. Battey⁴, Yanxiang Zhou^{1,5}, Rebecca C. Wade¹, Amit Banerjee^{6,7}

¹ EML Research gGmbH, Villa Bosch, Schloss-Wolfsbrunnengasse 33, 69118 Heidelberg, Germany, ² Department of Molecular and Genomic Bioinformatics, Denis Diderot University – Paris 7, Paris, France, ³ Computational Biochemistry, IWR, University of Heidelberg, 69120 Heidelberg, Germany, ⁴ Swiss Institute of Bioinformatics and Biozentrum, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland, ⁵ Department of Chemistry and Biochemistry, Ludwig-Maximilians-University Munich, Butenandtstrasse 5-13, 81377 Munich, Germany, ⁶ Department of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy and Health Sciences, Wayne State University, 259 Mack Avenue, Detroit MI 48201-2417, USA, ⁷ Department of International Medical Education and Development, Rheinische Friedrich-Wilhelms-Universität, Faculty of Medicine, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Molecular electrostatic potential as a predictor of protein function
 - 3.1. Examples of electrostatic potential as a predictor of function
 - 3.2. Calculating and comparing electrostatic potentials
 - 3.3. Electrostatic potentials and UFP ligation
 - 3.4. Electrostatically key residues
 - 3.5. Electrostatic potential and E2 function
 - 3.6. Other UFP related proteins and their electrostatic potentials
 - 3.7. The relationship between sequence and electrostatic potential
 - 3.8. A webserver of ubiquitin related structures and electrostatic potentials
 - 3.9. Ensuring the suitability of protein structures for calculating and comparing electrostatic potentials
4. Perspective
5. Acknowledgment
6. Reference

1. ABSTRACT

Post-translational modification by ubiquitin and ubiquitin-like (UBL) proteins is a key mechanism for cellular control. The specificity of the enzymes of ubiquitination and their close paralogs is dependent on their molecular electrostatic potentials. For example, analysis of molecular electrostatic potentials and electrostatically key residues can account for the selectivity of different E1s (activating enzymes) and of different SUMO proteases. The molecular interactions of the ubiquitin conjugating enzymes, the ubiquitin family proteins (UFP) and UBL domains are discussed in detail. An interesting observation is that the Non Canonical Ubiquitin Conjugating Enzymes (NCUBEs) have electrostatic potentials that are more similar to the UBC9 orthologs, the SUMO conjugating enzymes, than they are to other ubiquitin conjugating enzymes. It had previously been suggested that UBC9 may select for SUMO based on its difference in electrostatic potential as compared to other E2s but the NCUBE exception suggests that this may not be the case. The web site <http://www.ubiquitin-resource.org/> allows users to find the E2s most electrostatically similar to a query E2. Where possible, models have been made for all E2 domains in the SMART database (<http://smart.embl-heidelberg.de/>). A brief overview of molecular electrostatic potentials and their application to understanding protein function is also given.

2. INTRODUCTION

Ubiquitin family proteins (UFPs) regulate many cellular functions (1). Most commonly, the C-terminus of a UFP is ligated, via an isopeptide linkage, to the N(ε) of a lysine residue of the target protein. These post-translational changes are needed for many aspects of cellular regulation. Examples for which details are known include endocytosis, cell cycle progression, degradation of defective proteins, and DNA repair (1). Recently discovered are the MUBs, membrane-anchored ubiquitin fold proteins. These are prenylated, or otherwise post-translationally modified, such that they can bind to the membrane (2). At the time of writing, their physiological function is not clear, but it seems that they may provide a recognition site on a membrane for other proteins to bind to. As well as UFPs, there are also ubiquitin like domains (UBL domains) in proteins. These are involved in mediating protein-protein interactions, typically in a UFP ligation pathway or cell signalling (1).

Malfunction of the UFP ligation pathways has been associated with many diseases, including neurodegeneration and cancer (3). In particular, the UFP ligation pathways are often exploited by pathogens to enhance their infectivity, e.g. by exploiting the host cells' endocytotic machinery (4). Understanding the mechanisms by which UFP ligation occurs and is regulated will improve

the understanding of many diseases and may help therapeutic intervention. The basic mechanism of ubiquitination is known, but the details are not understood.

The basic mechanism by which most UFPs are targeted to their protein substrates is thought to be similar (5), although HUB1 and MUBs may prove to be exceptions (1, 2). An ATP dependent activating enzyme (E1) forms a thiol ester bond with the terminal glycine of the UFP. The UFP is transferred from the E1 to a UFP conjugating/carrier enzyme (UBC or E2), which then transfers the UFP to a target substrate, either directly, or with the help of a ligase enzyme (E3). In some cases, a further enzyme (E4) may also be required for the formation of multiubiquitin chains of length greater than four (6).

Given the similarities in UFP structure, and in their ligation pathways, how do the machineries of the different pathways recognize the different UFPs, the targets to which they should be ligated and the different ligation pathway proteins? It is possible to rationalise functional variation by amino acid sequence analysis, the postulation being that similar sequences have related evolutionary history and thus function. However, these analyses may mislead (7), particularly in predicting protein-protein interactions (8). Analysis of the physico-chemical properties of the structures of proteins, connects sequence to physics and chemistry, and thus to protein function. Compared to sequence analysis, such analyses can be time consuming but are often very informative (9-11). Analysis of the electrostatic potentials at the interfaces of six UFP ligation pathway complexes showed high levels of electrostatic complementarity at their surfaces (12). It has also been suggested that the molecular electrostatic potential of the SUMO conjugating E2 is important for its function (12-14), allowing it to distinguish SUMO from ubiquitin.

Here we discuss some general examples where protein function and electrostatic potential were found to be correlated. We then highlight examples from different UFP systems where electrostatic potential appears to be an important modulator of function. We review the high-throughput modeling of the E2s and the systematic analysis of their structures and their electrostatic potentials. This provides a complementary analysis to sequence and structure analyses, allowing further insights into E2 function. We also briefly describe www.ubiquitin-resource.org that allows searches for E2s of similar electrostatic potential, analogous to the sequence similarity searches that have been in common usage for many years. This website also includes additional analysis and the list of E2s is kept actual with the SMART database. Understanding the similarity between protein electrostatic potentials will enable a better understanding of protein interactions and functions. In particular, understanding the structural and functional determinants of the UFP ligation pathways will help the understanding of many disease states and should lead to better therapies, for example by the precise targeting of functionally important structural regions by structure-based drug design.

3. MOLECULAR ELECTROSTATIC POTENTIAL AS A PREDICTOR OF PROTEIN FUNCTION

3.1. Examples of electrostatic potential as a predictor of function

Electrostatic potentials are important mediators of protein-protein complex formation. Electrostatic potentials between interacting protein surfaces are expected to be complementary; although ions or small molecules trapped at a protein interface can bridge between residues that would be otherwise repulsive. Since a hydrophobic patch on a protein surface will tend to have a small electrostatic potential, compared to an acidic, basic or polar patch, the electrostatic potential should capture aspects of hydrophobic interaction properties. A study of the electrostatic potential of 12 protein-protein interfaces found anti-correlation (complementarity) between the electrostatic potentials of interacting proteins (15). The charges at the interfaces were not correlated, implying non-local effects are important in establishing electrostatic complementarity, e.g. charges distant from the interface and the shape of the dielectric boundary. Thus proteins with similar electrostatic potentials should interact with similar partners. An example of this is a non-catalytic role for acetylcholinesterase in the developing nervous system. The cholinesterases and structurally similar neural cell-adhesion proteins have a conserved electrostatic motif (16). Experimental work supports a role for this conserved electrostatic patch in interactions between acetylcholinesterase and the proteins laminin-1 and collagen IV (17). Rates of protein association and enzymatic rate constants have been predicted from electrostatic potentials (18, 19). Another good example showing the relationship between a protein's electrostatic potential and its function is barnase and its inhibitor barstar. They have highly complementary electrostatic potentials and a very high association rate. The experimental rates of association of different mutants can be correlated with the rate constants computed from Brownian dynamics simulations with electrostatic forces (20).

Electrostatic potentials are also important in predicting electron transfer between proteins. The physiologically equivalent but structurally divergent proteins cytochrome c6 and plastocyanin were superposed based on their electrostatic potentials (10). The aromatic residues tryptophan 63 in cytochrome c6 and tyrosine 83 in plastocyanin superposed. The position of tryptophan 63 is conserved across species as an aromatic residue. Thus, W63 and Y83 were inferred to have equivalent functions. Ferredoxin and (21) flavodoxin were similarly compared. Under conditions of iron deficiency, flavodoxin fulfills the function of the iron-sulfur protein ferredoxin, but they have different structures. Alignment of their electrostatic potentials overlays their active sites, rather than their centers of mass, in agreement with experimental data (9). The rate of electron transfer between plastocyanin mutants and cytochrome f correlate with the mutants' relative electrostatic similarities (22).

Protein electrostatic potentials are also often important for understanding how small chemicals interact

with proteins. The following example shows how evolution may conserve function without necessarily conserving sequence. Copper zinc superoxide dismutase (CuZnSOD) is a highly efficient enzyme. Its electrostatic potential guides the superoxide substrate to the active site, and enhancing this electrostatic guidance makes the enzyme even more efficient (23). Although residues identified as being important for the electrostatic potential of CuZnSOD in one species are not strictly conserved in other species, around the active site of CuZnSOD the electrostatic potential is conserved (11, 24). For the potential to be conserved a mutation of a charged residue must be accompanied by compensatory mutations elsewhere in the enzyme (11). For CuZnSOD, the evolution of charged residues mirrors the evolution of different species (11). For example, the prokaryotic enzymes have an additional loop (KDXK) compared to the eukaryotic enzymes, whereas other charged residues deemed important for eukaryotic CuZnSOD function are not present (11, 24).

There are many more examples in the literature of the importance of protein electrostatic potentials. For this reason, websites have been developed to serve the electrostatic potentials of the chains and complexes in the protein databank (<http://ef-site.hgc.jp/>) (25), to visualise the electrostatic properties at protein-protein interfaces (<http://projects.villa-bosch.de/dbase/molsurfer/>) (21), to compare the similarity of structures and their electrostatic potentials (<http://www.ubiquitin-resource.org>) (26), and to compare electrostatic potentials in a structure independent fashion (<http://agave.wustl.edu/similarity/>) (27). The focus of this article is the role of electrostatics in the ubiquitin and related systems.

3.2. Calculating and comparing electrostatic potentials

The electrostatic potential of a protein, evaluated at a given distance from the protein, could be obtained from the atomic charges of a standard molecular mechanics forcefield and Coulomb's law. However, this would ignore the effects of surrounding solvent and ions on a protein's potential. These effects can be included implicitly using the Poisson-Boltzmann equation. This equation and its broader application to bio-macromolecules is reviewed elsewhere (28). Since the aim here is to classify the similarities and differences of the electrostatic potentials of different proteins it is interesting to understand how different proteins behave at different ionic strengths.

Once the electrostatic potential of a protein is known, it is useful to compare it to that of other proteins. One way of doing this is Protein Interaction Property Similarity Analysis (PIPSA) (29, 30). In PIPSA, two similarity indices are used to quantify the similarity of electrostatic potentials, these are the Hodgkin index ($SI_{Hodgkin}$) and the Carbo index (SI_{Carbo}), eqs. 1 and 2. $SI_{Hodgkin}$ is +1 when the potentials are identical, and -1 if they are the exact inverse of each other. SI_{Carbo} is +1 if the two potentials are correlated, and -1 when they are anti-correlated; the relative magnitude of the two potentials does not affect the Carbo index. PIPSA compares the electrostatic potentials in 'skins' around the molecular surface. The skin is defined by a distance σ from the

molecular surface, and a skin thickness of δ . In the calculations presented here these parameters were 3 and 4 Å respectively. The similarity index may be converted to a distance (D), such that the proteins (a and b) that are most electrostatically similar (highest similarity index, SI_{ab}) have the shortest distance between them (eq 3).

$$SI_{Hodgkin} = \frac{2 \int \Phi_a \Phi_b dv}{\int \Phi_a^2 dv + \int \Phi_b^2 dv} \quad (1)$$

$$SI_{Carbo} = \frac{\int \Phi_a \Phi_b dv}{(\int \Phi_a^2 dv)^{1/2} (\int \Phi_b^2 dv)^{1/2}} \quad (2)$$

$$D = \sqrt{2 - 2SI_{ab}} \quad (3)$$

Comparison of protein electrostatic potentials is assisted by their visualisation. The potential at a distance from a protein is of interest, since it allows one to understand how two proteins may be attracted to each other, e.g. by alignment of their molecular electrostatic dipoles. In this case, isopotential surfaces are interesting since they indicate how the electrostatic potential varies through space. However, two proteins that might attract each other may not form a stable complex, because specific residue contacts on the two protein surfaces may not be complementary. In this case the electrostatic potential at a molecular surface may be more interesting.

3.3. Electrostatic potentials and UFP ligation

Ubiquitin and NEDD8 have highly similar electrostatic potentials (similarity of about 0.8, Table 1). NEDD8 is activated slowly by E1(ub), and A72R NEDD8 is activated almost as efficiently as ubiquitin (31). Ubiquitin R72L may also be activated by E1(NEDD8) and transthiolation to human UBC12. The importance of the mutation at position 72 is discussed in more detail later. The results are indicative that proteins with highly similar electrostatic potentials should have similar interaction properties.

SUMO and ubiquitin have anti-correlated electrostatic potentials. Smt3 (yeast SUMO) and human SUMO-1 have a Hodgkin similarity of about -0.5 with respect to ubiquitin. This is despite their having 40% sequence similarity, over the range of amino-acids compared here. Ubiquitination at a given lysine may be inhibited by sumoylation of the lysine (36), it is thus presumably important that SUMO and ubiquitin have very different interaction properties. There are proteins that have electrostatic potentials that are more anti-correlated with respect to ubiquitin, examples include 1EO6 (GATE-16) and the GABARAP proteins (similarity of IKLVA to 1UBQ is ~ -0.6). The most anticorrelated pair of UBLs in this study is 1NI2B and 2AX5A (respectively the FERM domain of ezrin from *E. coli* and URM1 from yeast), which has a Carbo similarity of -0.9 (Figure 1).

A non-covalent interaction between SUMO and UBC9 has been observed (37-39). The interacting surfaces

Table 1. The Hodgkin and Carbo molecular electrostatic potential similarity indices¹, and the percentage sequence similarity and identity² of selected UFPs and UBL domains

Protein name, species (PDB ID) ³ , number of residues compared	Ubiquitin, Human (1UBQ ⁴), 72	SUMO-1, Human (1A5R1), 73	SUMO-2, Human (1WM2A), 73	SUMO-3, Human (2D07B), 73	NEDD8, Human (1NDDC ²), 72
Ubiquitin, Bovine (1AARA ^{4,5}), 72	0.88 0.88 100 100	-0.62 -0.64 40 16	-0.41 -0.45 37 13	-0.32 -0.34 36 13	0.74 0.77 76 55
NEDD8, <i>A. thaliana</i> (1BT0), 72	0.80 0.81 79 61	-0.42 -0.43 50 23	-0.29 -0.31 37 15	-0.18 -0.19 36 15	0.82 0.83 91 84
E1(NEDD8), Human (1Y8XB), 92	0.41 0.43 26 9	-0.07 -0.08 24 12	-0.23 -0.28 23 10	-0.19 -0.22 23 11	0.48 0.58 27 11
UBL3, <i>M. musculus</i> (1WGH), 83	0.56 0.58 43 26	-0.40 -0.40 27 16	-0.64 -0.67 30 16	-0.62 -0.64 29 16	0.72 0.73 37 25
SMT3, <i>S. cerevisiae</i> (1L2NA), 72	-0.42 -0.46 34 11	0.77 0.75 75 51	0.70 0.70 62 47	0.62 0.62 61 47	-0.48 -0.49 37 19
SUMO-1, Human (1A5R1), 73	-0.49 -0.53 40 16	1.0 1.0 100 100	0.70 0.70 69 50	0.59 0.59 69 51	-0.40 -0.40 43 19
SUMO-2, Human (1WM3A ⁴), 72	-0.37 -0.42 36 13	0.64 0.65 69 51	0.98 0.98 100 100	0.97 0.97 100 100	-0.55 -0.56 38 16
SUMO-3, Human (2D07B), 73	-0.29 -0.32 37 13	0.59 0.59 69 50	0.95 0.95 100 100	1.0 1.0 100 100	-0.53 -0.54 38 16
E1(SUMO), Human, (1Y8QD), 90	-0.35 -0.40 29 16	0.16 0.16 26 9	0.19 0.19 21 6	0.04 0.05 22 6	-0.33 -0.34 33 15
GATE-16, Bovine, (1E06A), 85	-0.56 -0.58 19 8	0.58 0.58 21 8	0.66 0.68 28 12	0.59 0.59 27 12	-0.62 -0.62 16 6
GABARAP, Human, (1GNUM), 85	-0.48 -0.51 16 8	0.50 0.50 23 12	0.65 0.66 26 10	0.62 0.63 25 11	-0.70 -0.70 12 4
FERM, domain, <i>E. coli</i> (1NI2A), 80	-0.20 -0.21 27 11	0.40 0.40 27 12	0.11 0.11 26 12	0.06 0.06 25 12	-0.12 -0.12 34 15

¹ $SI_{Hodgkin}$ and SI_{Carbo} are the first and second number in the main cells, respectively, and are shown in standard typeface. ² third and fourth number in each cell are, respectively, the sequence similarity (using the BLOSUM 90 matrix), and the sequence identity, shown in italics. ³ PDB IDs include chain identifier or the NMR model. ⁴ These structures were slightly shorter than the other structures due to residues not being resolved in the experimentally published structure, or in the case of ubiquitin due to it having a shorter N-terminus than the other structures. The procedure for preparing structures is discussed in the caption of (Figure 1). ⁵ monomer from diubiquitin.

have highly complementary electrostatic potentials (37). The interacting surface is highly conserved in SUMO-1, SUMO-2 and SUMO-3 and all three were found to make the same NMR footprint on UBC9 (38). The binding site on UBC9 of E1(SUMO) overlaps with the non-covalent interaction site of SUMO, both in human (38) and yeast (39). However, the UBL domain of E1(SUMO) does not overlap with the non-covalent interaction site of SUMO (40). The structure of the E1(SUMO) UBL domain bound to UBC9 has not been solved, but the structure of the UBL domain of E1(NEDD8) bound to UBC12 has been solved (41). In the left view of 1Y8XB (Figure 1), UBC12 would bind to the top left corner of the UBL domain. In the UBL domain of E1(SUMO) this corresponds to the negative lobe of electrostatic potential (Figure 1, 1Y8QD), suggesting that UBC9 will bind to it in a similar fashion, the N-terminal helix of UBC9 having a positive electrostatic potential. In contrast, SUMO binds non-covalently to UBC9 using its beta-sheet face (37). Thus, despite the similarity of their protein folds, SUMO and the UBL domain of E1(SUMO) probably bind UBC9 using different surfaces.

The biological function of a non-covalent interaction between SUMO and UBC9 is not clear. UBCH5c appears to form non-covalent interactions with the ubiquitin thiol-ester linked to a neighbouring UBCH5c. These may be important for processivity, i.e. they allow very rapid formation of ubiquitin chains on BRCA1. A non-covalent interaction between ubiquitin and human UBC2b has also been seen (42), although no functional role has been proposed in this case. Non-covalent interactions between ubiquitin and the E2 core

are not generally observed (43, 44), although the UBA domain of UBC1 binds ubiquitin and may have a role in polyubiquitin chain synthesis (45).

3.4. Electrostatically key residues

The specificity of NEDD8 or ubiquitin for their appropriate ligation pathways is controlled by residue 72 at the stage of E1 binding (46). This residue is alanine in NEDD8, arginine in ubiquitin and glutamate or glutamine in SUMO orthologs, and is complementary to residues on the respective E1 (L206/Y207 in E1(NEDD8), D in E1(U), K in E1(S)) (46). Yeast SUMO (Smt3) can be activated by human E1(SUMO), transferred to human UBC9 (E2I) and conjugated to RanGAP1, although less efficiently than the human SUMO (40). A more complicated case is ISG15. ISG15 is ligated like ubiquitin to other proteins (47), and is thought to be conjugated by UBCH8/E2L6 (48), an E2 that also conjugates ubiquitin. However, ISG15 has its own E1, UBE1L (49). ISG15 is a UFP with two UBL domains, the N-terminal domain having very little electrostatic similarity to SUMO, but none to ubiquitin ($SI_{Hodgkin}$ of 0.3 and -0.1 respectively), the C-terminal domain being much more electrostatically similar to ubiquitin than SUMO ($SI_{Hodgkin}$ of 0.4 and -0.5 respectively). Mature ISG15 has a C-terminus that is almost identical to ubiquitin, including conservation of R72 (VLR⁷²LRLRGG in ubiquitin, compared to NLRLRLRGG in ISG15) (47, 50). Since ISG15 has its own E1 and presumably does not bind E1(ubiquitin), it is clear that R72 alone is not enough to allow selection for E1(ubiquitin). In the case of NEDD8 and ubiquitin, they have similar enough molecular surfaces that the identity of residue 72 becomes decisive.

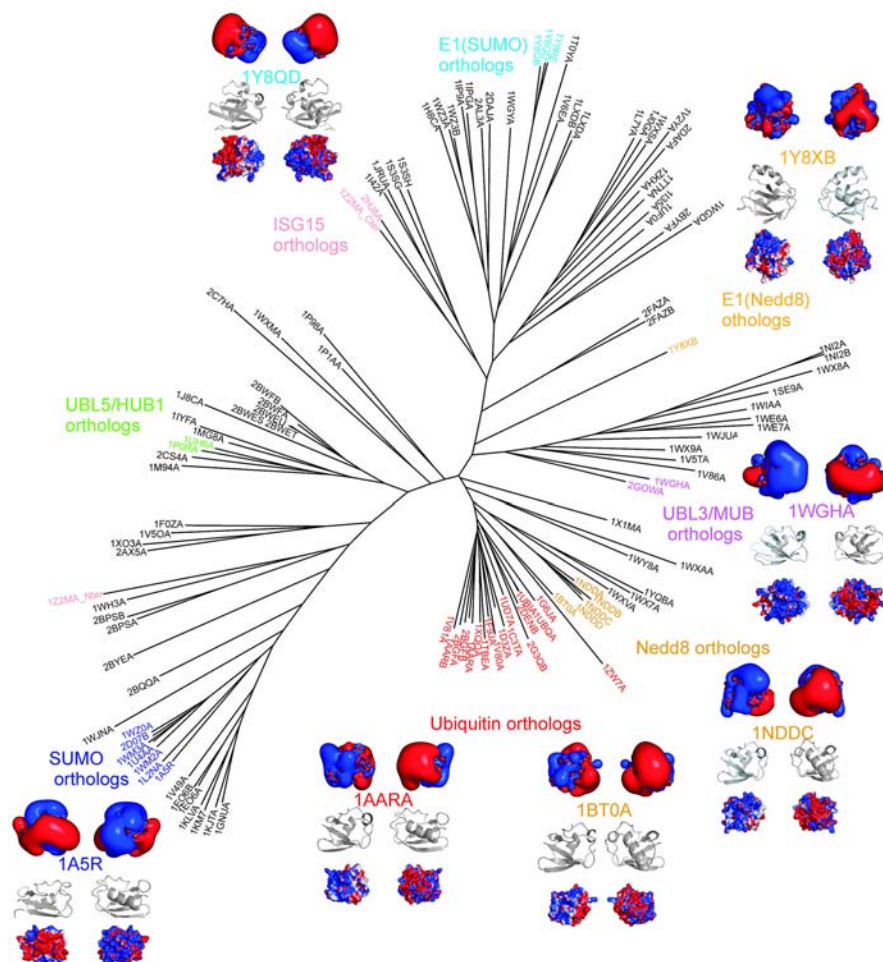


Figure 1. Epogram of ubiquitin family proteins and ubiquitin-like protein domains, showing the similarity between their molecular electrostatic potentials. Each leaf of the dendrogram is labelled with the corresponding protein's protein data bank code (www.rcsb.org) with the fifth character being the chain identifier. The dendrogram was constructed with a Neighbor-Joining algorithm (32) using distances calculated from SI_{Carbo} as defined in eq 3. The branch length between two leaves on the plot is proportional to the calculated distances when the leaves are close. However, the clustering process is such that this relationship is compromised as leaves become further apart. The points furthest apart in the plot are thus not necessarily the most dissimilar. ISG15 is a UFP consisting of two UBL domains; the C- and N-terminal domains of 1Z2M are indicated, colored pink, in different parts of the Epogram as 1Z2MA_Cter and 1Z2MA_Nter respectively. Two views of representative potentials are shown close to their corresponding entries in the Epogram. These were rendered with PyMOL (33). The views are related by a 180 degree rotation around the vertical axis. Isopotential contours are at $+0.5$ kT/e (blue) and -0.5 kT/e (red). Surface potentials are color ramped from red (-1 kT/e) to blue ($+1$ kT/e) via white (neutral). An ionic strength of 50mM and temperature of 300K were used in the calculations, which were performed with UHBD (34). We found 57 UFPs in the PDB and 58 UBL domains from multidomain proteins (7th August 2006). Each structure was fit onto pdb id 1A5R, using the program FAST (35). 1A5R was truncated to include residues from Y21 to E93. Residues beyond this range were highly variable in structure, and thus comparison would not be meaningful. All other UBL structures were truncated to have their N- and C-termini structurally equivalent to those in 1A5R. Other details of the calculation and figure production are given in (12). The following proteins are represented in the plot: 1GNU Human GABARAP; 1KJT Rat GABARAP; 1KM7 Human GABARAP; 1KLV Human GABARAP; 1EO6 bovine GATE-16; 1V49 Human microtubule associated protein; 1A5R human SUMO-1; 1L2N Smt3; 1WM2 human SUMO-2; 1U41 human SUMO-3 C47S; 1WM3 Human SUMO-2; 2D07 Human SUMO-3; 1WZ0 Human SUMO-2; 1WJN Rat mouse C-terminal UBL of tubulin-specific chaperone E; 2BQQ human N-terminal domain of doublecortin; 2BYE human phospholipase c epsilon Ras association domain; 2BPS *Bacillus subtilis* UBL of YUKD; 1WH3 human C-terminal UBL of protein; 1Z2M human ISG15; 2AX5 yeast URM1; 1XO3 mouse UBL; 1V50 mouse UBL domain; 1F0ZA *E. coli* this; 1M94 *S. cerevisiae* HUB1; 2CS4 Human Ras association domain; 1P0R Human UBL5; 1UH6 Mouse UBL5; 1MG8 Mouse UBL domain of parkin; 1IYF Human UBL domain of parkin; 1J8C Human UBL domain of hPLIC2; 2BWE *S. cerevisiae* UBL domain of DSK2; 2BWF *S. cerevisiae* UBL domain of DSK2; 2C7H Human DWNN domain of RBBP6; 1WXM Human Ras-binding domain in A-Raf kinase; 1P1A Human UBL domain of hHR23B; 1P98 Human UBL domain of hHR23B; 2HJ8 Human C-terminal domain of ISG15; 1I42 Rat UBX domain of P47; 1JRU Rat UBX domain of P47; 1S3S Mouse P47; 1H8C Human UBX domain of FAF1; 1WZ3 *A. thaliana* ATG12; 1IP9 *S. cerevisiae* PB1 domain of Bem1p; 1IPG *S. cerevisiae* PB1 domain of Bem1p; 2AL3 Mouse UBL domain of TUG; 2DAJ Human UBL domain in COBL-like 1 protein; 1WGY Human Ras/Rap1-associating domain of RAP1; 1Y8Q Human UBL of E1(SUMO); 1Y8R Human UBL of E1(SUMO); 1TOY *C. elegans* UBL domain of tubulin binding cofactor B; 1V6E Mouse UBL domain of tubulin-specific chaperone B; 1LXD Rat Ras interacting domain of RalGDS; 1L7Y *C. elegans* ZK652.3; 1J0G Mouse hypothetical proteins; 1WXS Human UFM1; 1V2Y Mouse UBL fold; 2DAF Human UBL domain of hypothetical protein; 1ZKH Human UBL domain in SF3A1; 1TTN Human UBL domain of DC-UBP; 1I35 *S. Pombe* Ras-binding domain of Byr2; 1UF0 Human DCX domain of doublecortin-like kinase; 2BYF Human Ras associated domain of phospholipase C; 1WGD Human UBL domain of Herp; 2FAZ Human UBL domain of nuclear zinc finger protein; 1Y8X Human UBL of E1(Nedd8); 1NI2 *E. coli* FERM domain of ezrin; 1WX8 Mouse UBL domain; 1SE9 *A. thaliana* UBL protein; 1WIA Mouse UBL; 1WE6 *A. thaliana* UBL domain in splicing factor; 1WE7 Mouse UBL domain splicing factor SF3a120; 1WJU Human UBL domain of NUB1; 1WX9 Human UBL domain of BAT3; 1V5T Mouse UBL domain; 1V86 Mouse UBL domain; 1WGH Mouse MUB/UBL3; 2GOW Human MUB/UBL3; 1X1M Mouse UBL domain in SB132; 1WXA Mouse Ras-binding domain; 1WY8 Human UBL domain of NP95/ICBP90-like ring finger protein; 1YQB Human ubiquitin 3; 1WX7 Human ubiquitin 3; 1WXV Human UBL domain of BCL-2 binding anthanogene-1; 1NDD Human NEDD8; 1BT0 *A. thaliana* RUB1; 1ZW7 *S. cerevisiae* ubiquitin E34P; 1G6J Human ubiquitin; 1UBI Human ubiquitin; 1UBQ Human ubiquitin; 2DEN Human ubiquitin; 2G3Q *S. cerevisiae* ubiquitin; 1UD7 engineered ubiquitin I3V V5L L151 V26F I67L; 1C3T Mouse ubiquitin; 1D3Z human ubiquitin; 1V80 bovine ubiquitin; 1F9J human ubiquitin (from tetraubiquitin); 1TBE human ubiquitin (from tetraubiquitin); 1XQQ human ubiquitin; 1AAR – bovine ubiquitin; 2BGF human ubiquitin (from di-ubiquitin); 1V81 bovine ubiquitin.

A similar identification mechanism is used in SUMO maturation (51, 52). Before newly synthesised SUMO can be ligated to its target, its most C-terminal residues must be cleaved off. This maturation process is performed by SUMO specific proteases; these also deSUMOylate SUMOylated proteins as part of the SUMO regulation pathway. The SUMO specific protease SENP1 promotes the maturation of SUMO-1 in preference to SUMO-2 (52). However, it removes SUMO-1 and SUMO-2 from their target proteins with almost equal processivity. SENP1 has an acidic negatively charged surface patch at the point where it binds the SUMO C-terminal peptide. SENP-2, which preferably matures SUMO-2, is neutral at the same position. The amino acids in SUMO-1 and SUMO-2 that determine their specificity are respectively HS and VP (52).

3.5. Electrostatic potential and E2 function

Cell free assays indicate that E2C and sequentially similar proteins ubiquitinate mitotic cyclins (53-56), thus leading to exit from mitosis (57). Yeast UBC11 (ScUBC11), the E2 from *S. cerevisiae* that is most similar to E2C, is not essential and does not ubiquitinate cyclin B in a clam derived cell-free system (58). Compared to the other E2C orthologs, ScUBC11 lacks a 22 amino acid N-terminal extension. This may explain the different behaviour (58), however the extension is not well conserved in the other E2C orthologs. The electrostatic potential of ScUBC11 is also very different from its sequence orthologs (Figure 2, Figure 3 and (12)). ScUBC11 is the most negative of the cluster of E2C sequence orthologs (Human E2C, clam E2C, *S. pombe* UBC11, and ScUBC11). At 0mM ionic strength, the other sequence orthologs of E2C still cluster well, but ScUBC11 is more similar to yeast UBC2 and UBC10 than it is to human or clam E2C (Figure 3a and (12)). Whilst ScUBC11 is more similar to *S. pombe* UBC11 ($SI_{Hodgkin}$ 0.81, SI_{Carbo} 0.87) than to yeast UBC2 ($SI_{Hodgkin}$ and SI_{Carbo} of 0.80), it has much less similarity to human or clam E2C ($SI_{Hodgkin}$ to clam E2C 0.67, SI_{Carbo} 0.74) and so is not clustered with them. *S. pombe* UBC11 has a similarity to clam E2C of 0.84, hence it clusters with the other two E2C orthologs in this plot. As the ionic strength in the Poisson-Boltzmann calculations increases, ScUBC11 becomes much less similar to yeast UBC2 but only slightly less similar to the E2C orthologs (50mM, $SI_{Hodgkin}$, yeast UBC2:0.61, *S. pombe* UBC11: 0.72, clam E2C:0.66; 150mM, yeast UBC2:0.52, *S. pombe* UBC11: 0.67, clam E2C:0.64). Given the difference in the electrostatic properties of ScUBC11 compared to clam E2C, it is thus unlikely to interact with the same proteins as clam E2C. The other orthologs ubiquitinate mitotic cyclins in a clam cell-free system. Based on its electrostatic potential, ScUBC11 is unlikely to interact with the same proteins as its orthologs, and so is not expected to interact with the clam machinery necessary for cyclin B ubiquitination. However, the yeast machinery necessary for mitotic cyclin ubiquitination may have evolved a different electrostatic potential from that in clam. Thus, ScUBC11 may still be able to interact with the yeast

machinery. Knockouts of ScUBC11 are viable (58) suggesting that another E2 may take over the ScUBC11 role once it has been knocked out. A C93S ScUBC11 mutant, expected to show dominant negative behaviour, is reported to have no effect on cell growth, but over expression of wild type ScUBC11 was inhibitory to cell growth (58). No other yeast E2 has been found to be solely responsible for cyclin B ubiquitination and no other role for ScUBC11 has been proposed.

The last paragraph illustrates how high-throughput comparison of protein electrostatic potentials can be useful for experimental design and interpretation. Proteins that are likely to have similar behaviour may be selected for an assay. Proteins that are likely to differ functionally from known proteins may also be selected, depending on what is sought.

The yeast proteins Rad6/UBC2, Cdc34/UBC3, UBC8 and human E2H have all been reported to ubiquitinate histones (H2A or H2B) *in vitro* without requiring an E3 (59-63), although the UBC8 result is contradicted elsewhere (64). All four of these proteins have similar strongly negative electrostatic potentials (Figure 2, Figure 3, and (12)) and acidic C-terminal extensions. Indeed, the ability of Cdc34, Rad6 and E2H to ubiquitinate histones *in vitro* has been shown to be dependent on the presence of their C-terminal tail (59, 62, 65). This is presumably related to them being attracted to the positive potential of the histones. In *S. cerevisiae*, only Rad6 is responsible for histone ubiquitination (63).

Whilst the yeast E2s have predominantly negative electrostatic potentials, a subset have very positive potentials (fig 2). The most positive are the SUMO conjugators (UBC9, E2I), and this has been proposed as a mechanism for them to distinguish SUMO from ubiquitin (12-14, 37). Since ubiquitin and SUMO both have a negative and a positive face, albeit anticorrelated (Figure 1), this cannot be the sole mechanism of discrimination. In addition to the UBC9 orthologs, the NCUBE proteins (UBC6, E2J2, E2E2) have strongly negative potentials, notably along more of the N-terminal helix than is typical for the other E2s. Nonetheless, these E2s are thought to conjugate ubiquitin. The positive potential may be important for target protein recognition by the E2 (13).

E2L6 has a distinct electrostatic potential compared to the other E2s. It is worth remarking on this since it is responsible for ISG15 conjugation and ubiquitin conjugation (48). It was previously suggested that uncharacterised E2s with unusual electrostatic properties may also have unusual and novel functions (12), and this example would appear to support this.

UBC12 and E2M conjugate NEDD8, in yeast and human respectively. However these two proteins have different electrostatic potentials (Figure 2, Figure 3, a Hodgkin similarity of 0.6 at 50mM and 150mM, 0.7 at 0mM). This suggests that they may not efficiently substitute for each other.

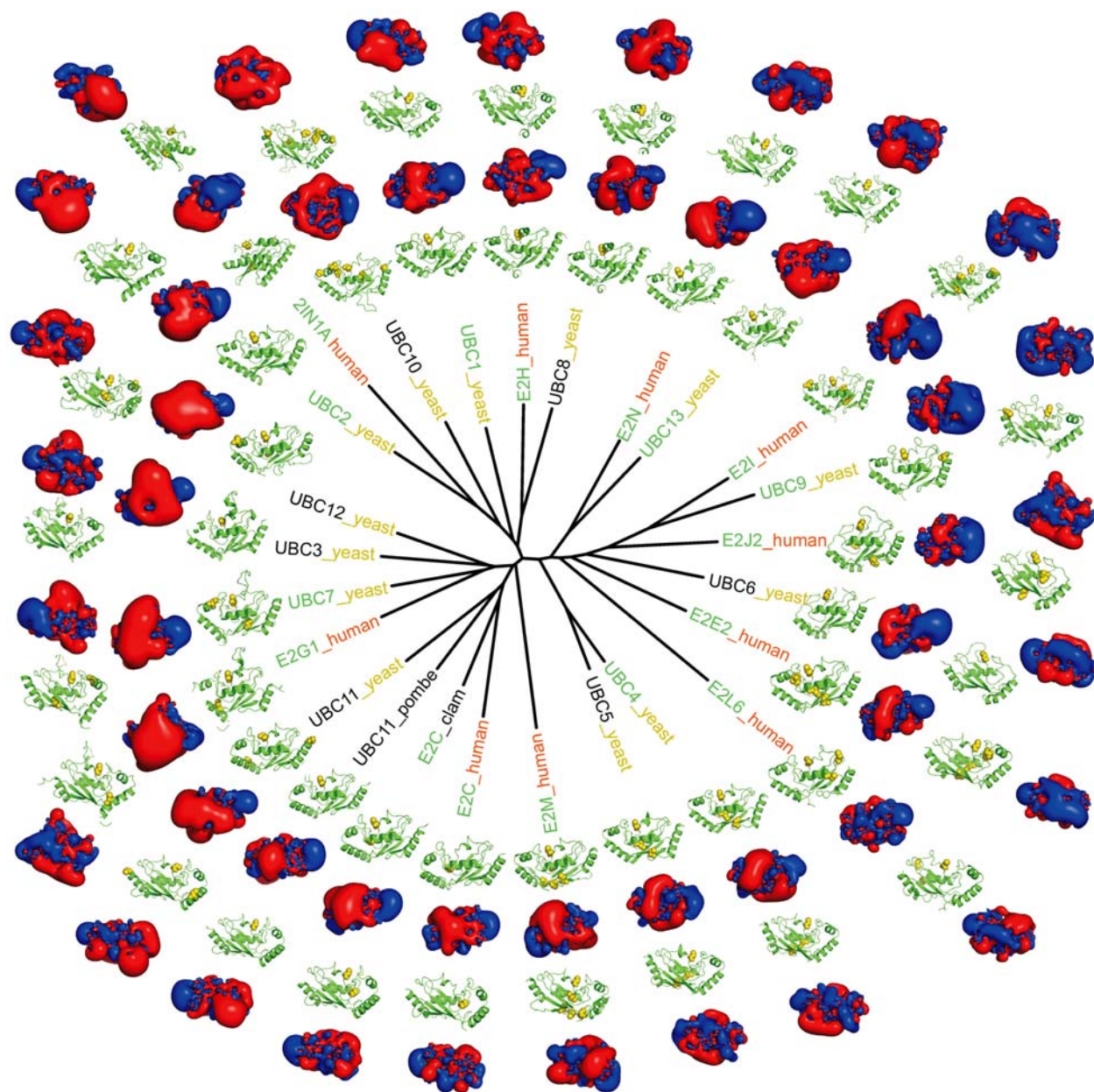


Figure 2. Epogram of yeast and selected other E2 enzymes. The distance between two dendrogram leaves is related to the SI_{Carbo} as discussed in the caption to (Figure 1). The structures that were analysed were those E2s experimentally determined for *S. cerevisiae*, comparative models of those yeast E2s not determined experimentally, and the templates used for the homology modeling. The structures with the following pdb codes or sequences with the following Swissprot codes were used: UBC1(yeast), 1FZY; UBC2/RAD6(yeast), 1AYZ; UBC3/CDC34(yeast), P14682; UBC4(yeast), 1QCQ; UBC5(yeast), P15732; UBC6(yeast), P33296; UBC7(yeast), 2UCZ; UBC8(yeast), P28263; UBC9(yeast), 2GJD; UBC10(yeast), P29340; UBC11(yeast), P52492; UBC12(yeast), P52491; UBC13(yeast), 1JAT; E2J2(human), 2F4W; E2G1(human), 2AWF; E2H(human), 1YH6; E2L6(human), also known as UBC8, the E2 that carries ISG15, 1WZV; E2I(human), 1U9B; E2E2(human), 1Y6L; E2C(human), 1I7K; E2M(human), 1Y8X; E2N(human), 1J7D; human UFM1-conjugating enzyme, 2IN1; E2C(clam), 2E2C; UBC11(*S. pombe*), O00103. Calculations were performed and isopotentials displayed as described in the caption for (Figure 1), with two representations related by a 180 degree rotation about the vertical axis. The ribbon representations of the structures are colored green, with all cysteine residues represented as yellow spheres. The structure 1I7K (human E2C) is a C114S mutant, so there is no active site cysteine to be shown. Text on the leaves of the Epogram is colored green for crystal structures, red for human, yellow for *S. cerevisiae*, and otherwise black.

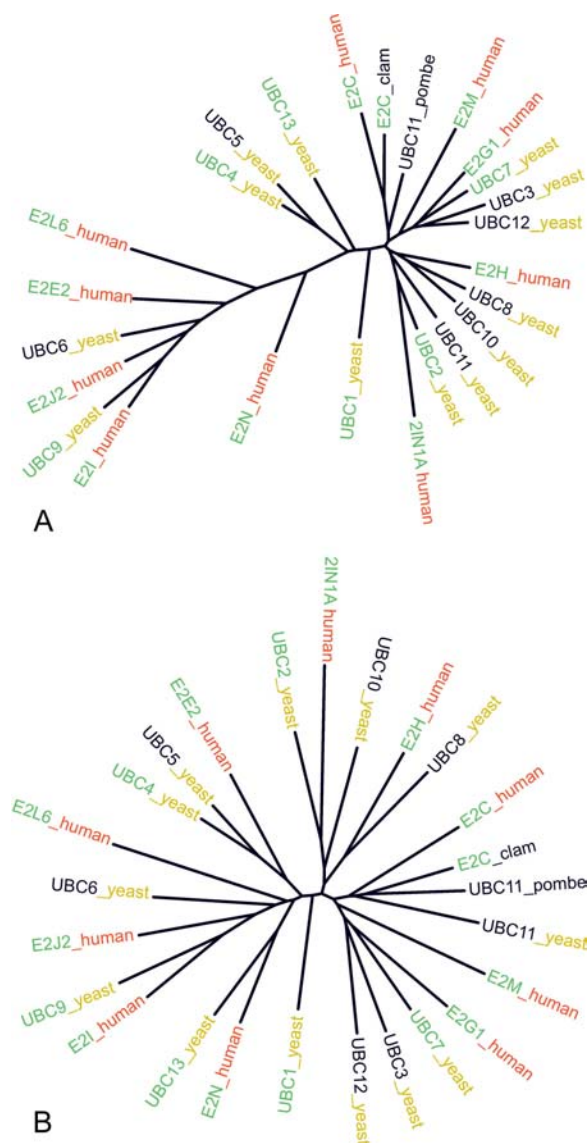


Figure 3. Epogram of yeast and selected other E2s at 0mM and 150 mM ionic strength. The Epograms are as described in (Figure 2), but with the Poisson-Boltzmann equation solved with ionic strength (a) 0mM (b) 150mM.

3.6. Other UFP related proteins and their electrostatic potentials

UBA domains recognize ubiquitin. Lowe *et al.* (66) discussed the role of the electrostatic potential of the UBA domain of Dsk2. They suggest that the potential is critical for the oligomerisation of multiple UBA domains into a helical structure that can recognize multi-K48-ubiquitin chains. The charged UBA residues important for the oligomerisation are not conserved in other UBA domains, suggesting that electrostatic potential may also be used by UBA domains to differentiate one from another.

Hydrolysis of ubiquitin ethyl ester by ubiquitin carboxyl-terminal hydrolase L3 (UCH-L3) is inhibited by increasing ionic strength (67). Since the structure of the

enzyme was not greatly perturbed by the increased ionic strength, it seems likely that the attraction between UCH-L3 and ubiquitin is largely electrostatic in nature.

3.7. The relationship between sequence and electrostatic potential

Comparative electrostatic potential analysis of the E2s gives different results to those of sequence analysis (12). This is most evident in the family of E2C proteins, discussed above. A look at Table 1 also shows the lack of correlation between sequence similarity and electrostatic potential, e.g. 1WGHA v 1UBQ has Hodgkin similarity of 0.56 and sequence similarity of 43%, compared to 1A5R1 and 1UBQ which have -0.49 and 40%, and 1WGHA and 1NDDC which have 0.72 and 37%. Elsewhere, little correlation was found between pairwise amino acid sequence similarity and pairwise electrostatic potential similarity (30, 68). To understand this, consider some hypothetical protein; mutating a few positively charged residues to negatively charged residues would create a large change in electrostatic potential but a small change in sequence similarity. Thus, it is easy to see why there is often no strong correlation between sequence identity and electrostatic potential. Indeed, sequence may evolve whilst electrostatic potential remains conserved, as discussed in the introduction to section 3. Equally, correlation between sequence and electrostatic potential has been found in orthologous proteins (69).

It is not clear when evolution will tend to conserve electrostatic potentials. Proteins that interact with cellular membranes or DNA have an evolutionary constraint due to the conserved charged nature of these surfaces. Similarly certain physical processes such as electron transfer or enzymatic catalysis may also impose evolutionary constraints on the electrostatic potential of a protein. However, for the general case of two or more interacting proteins, it is possible for correlated mutations to occur, and for their electrostatic potentials to mutate and thus be different between different species.

3.8. A webserver of ubiquitin related structures and electrostatic potentials

To comprehensively investigate the E2 proteins, we have developed a computational pipeline to model their structures and electrostatic potentials. E2 sequences are taken from the SMART domain database (70) as it is updated. Our current protocol extracts and models only one E2 domain per protein sequence. The template structures for the homology modelling are truncated two amino acids beyond the last secondary structure elements bounding the E2 domain. The target sequences, i.e. the sequences to be modelled, are truncated to match. The target sequences are aligned onto a set of sequence profiles and modelled by comparison with known structures. The template structures need pre-refinement as discussed elsewhere (26) and briefly in section 3.9. The electrostatic potentials are calculated using the linearised Poisson-Boltzmann equation (34).

The data that we have generated may be queried with Uniprot and Ensembl codes via the Website

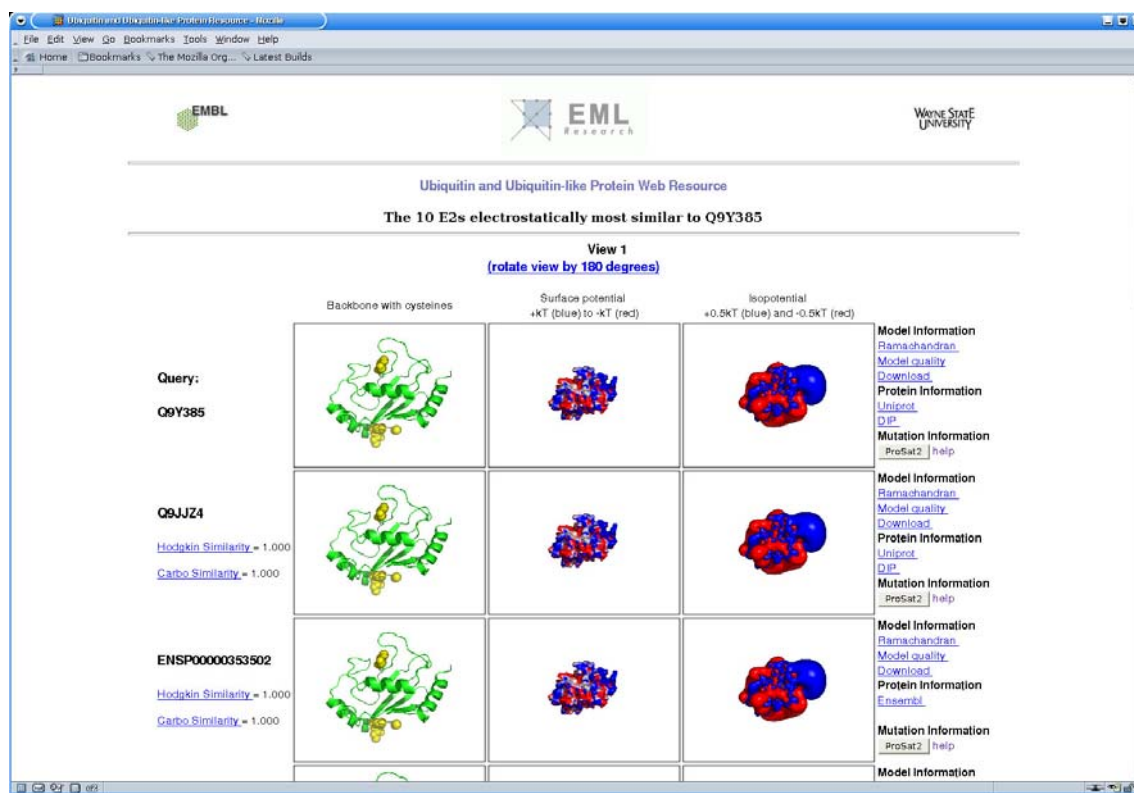


Figure 4. A snapshot of a search at <http://www.ubiquitin-resource.org> for the ten E2s with highest electrostatic potential similarity to a query protein, Uniprot code Q9Y385. A cartoon of the related structures, their potential on the solvent accessible surface and their isopotentials are returned. Each entry also has links to other potentially relevant data such as protein interactions (DIP) (72) and mutation data (ProSAT 2) (71). The potentials were calculated by solution of the linearised Poisson-Boltzmann equation with ionic strength 50mM. The isopotential contours are at ± 0.5 kT/e, the solvent accessible surface representation is coloured from red ($-kT/e$) to white (neutral) to blue ($+kT/e$).

<http://www.ubiquitin-resource.org>. The website can return the E2s that are electrostatically most similar to a query, or allow a comparison of a list of E2s. The results page displays pictures of the electrostatic potentials of the E2s, ribbon representations of their backbones, and the similarity indices of the protein pairs (Figure 4). The user may download the models of the E2 structures and the UHBD format grid file of the electrostatic potential, for displaying locally. A link to ProSAT 2 (71) allows amino-acid mutation data to be retrieved from databases such as Swissprot (www.expasy.org) and displayed on the modelled structures. There are also links to a sequence's Uniprot or Ensembl entry, to various data on the model quality and to the 'Database of Interacting Proteins' (DIP) (72). The website gives researchers in the field of ubiquitination easy access to the calculated electrostatic potentials. It allows them to compare potentials in the same way that they currently compare sequences, thus allowing a better understanding of how electrostatic potentials are related to function and to evolution.

3.9. Ensuring the suitability of protein structures for calculating and comparing electrostatic potentials

Care is required when performing calculations for the comparison of electrostatic potentials, as discussed in detail elsewhere (19, 26). Enzyme kinetic constants may be

predicted based on relative electrostatic similarity (19). The predictions may be poor if the compared structures have inconsistent sidechain rotamers for the residues around the active site, e.g. if a comparative model has different rotamers compared to identical residues in the template. The analysis of the UFPs (Figure 1) shows that whilst different crystal structures of e.g. ubiquitin have slightly different electrostatic potentials, they are still similar enough to cluster and thus allow quantitative interpretation. Similarly, analysis of the E2 enzymes showed homology models and crystal structures having similar electrostatic potentials (26). These results suggest that rotamer variation is not an immediate problem when one is interested in qualitative analysis. Nonetheless, care is needed when considering e.g. histidine rotamers as they may have different preferred protonation states (26). For qualitative comparison, small misalignments during the homology modelling have also been shown to be tolerable (12), although clearly not desirable. Also cause for consideration are variations in loop conformations between experimental structures and, for comparative modelling, unexpected structural elements compared to the template.

Inconsistencies and errors in the models will make it difficult to meaningfully compare the electrostatic potentials of different structures. In the www.ubiquitin-

resource.org pipeline, checks are made on both the sequences and the templates used in the modelling process (12, 26). For fully automated modelling, further checks are made on the resulting models. The sequence alignment of model and template are compared to determine if their N- and C-termini are congruent, i.e. does the model have termini as long as the template. An incomplete model will not give electrostatic potentials useful for comparison with other structures and is thus not included in the comparison of electrostatic potentials, although it may be useful for structural interpretation. Similarly, models with an insertion or deletion of more than five amino acids, ones that have unknown amino acid types (an X in the sequence entry), or ones that have abnormal structural characteristics (26, 73) are also not used for electrostatic calculations. Suitable care in the modelling and comparison process has allowed meaningful results, as discussed above and elsewhere (12, 19, 26).

4. PERSPECTIVE

The electrostatic potential of a protein is dependent on the three dimensional arrangement of its amino acids. The correlation between the linear amino acid sequence and electrostatic potential is weak. The interactions that a protein needs to function may place constraints on the biologically permissible electrostatic potential. These constraints can be investigated by comparison of the electrostatic properties of orthologous and paralogous proteins. The comparative study of electrostatic potentials provides information additional to that of the linear sequence and provides a different perspective for understanding functional and evolutionary variations and constraints.

Comparison of electrostatic potentials is more complex than sequence analysis, but the examples given here show that valuable insight can be gained. The UFP ligation pathways are interesting because of their functional diversity and biological importance. They are a good system for developing understanding of the relationship between electrostatic potential, evolution and function. High throughput analysis of electrostatic potentials, as presented here, will help to decipher the function and mechanism of different parts of the UFP systems. A long term goal is to be able to compare the electrostatic potentials of proteins as routinely as their sequences.

5. ACKNOWLEDGEMENTS

This work was supported by the Klaus Tschira Foundation and the National Institute of Health (USPHS) grant GM59467.

6. REFERENCES

1. R. L. Welchman, C. Gordon and R. J. Mayer: Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol* 6, 599-609 (2005)
2. B. P. Downes, S. A. Saracco, S. S. Lee, D. N. Crowell and R. D. Vierstra: MUBs, a family of ubiquitin-fold proteins that are plasma membrane-anchored by prenylation. *J Biol Chem* 281, 27145-57 (2006)

3. M. Reboud-Ravaux, Protein Degradation in Health and Disease Springer Verlag, Heidelberg. (2002)
4. D. Mukhopadhyay and H. Riezman: Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315, 201-5 (2007)
5. S. Goettsch and P. Bayer: Structural attributes in the conjugation of ubiquitin, SUMO and RUB to protein substrates. *Front Biosci* 7, a148-62 (2002)
6. M. Koegl, T. Hoppe, S. Schlenker, H. D. Ulrich, T. U. Mayer and S. Jentsch: A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 96, 635-44 (1999)
7. J. A. Gerlt and P. C. Babbitt: Can sequence determine function? *Genome Biol* 1, REVIEWS0005 (2000)
8. S. Mika and B. Rost: Protein-protein interactions more conserved within species than across species. *PLoS Comput Biol* 2, e79 (2006)
9. G. M. Ullmann, M. Hauswald, A. Jensen and E. W. Knapp: Structural alignment of ferredoxin and flavodoxin based on electrostatic potentials: implications for their interactions with photosystem I and ferredoxin-NADP reductase. *Proteins* 38, 301-9 (2000)
10. G. M. Ullmann, M. Hauswald, A. Jensen, N. M. Kostic and E. W. Knapp: Comparison of the physiologically equivalent proteins cytochrome c6 and plastocyanin on the basis of their electrostatic potentials. Tryptophan 63 in cytochrome c6 may be isofunctional with tyrosine 83 in plastocyanin. *Biochemistry* 36, 16187-96 (1997)
11. D. R. Livesay, P. Jambeck, A. Rojnuckarin and S. Subramaniam: Conservation of electrostatic properties within enzyme families and superfamilies. *Biochemistry* 42, 3464-73 (2003)
12. P. J. Winn, T. L. Religa, J. N. D. Battey, A. Banerjee and R. C. Wade: Determinants of functionality in the ubiquitin conjugating enzyme family. *Structure* 12, 1563-74 (2004)
13. H. Tong, G. Hateboer, A. Perrakis, R. Bernards and T. K. Sixma: Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *J Biol Chem* 272, 21381-7 (1997)
14. M. F. Giraud, J. M. Desterro and J. H. Naismith: Structure of ubiquitin-conjugating enzyme 9 displays significant differences with other ubiquitin-conjugating enzymes which may reflect its specificity for sumo rather than ubiquitin. *Acta Crystallogr D Biol Crystallogr* 54, 891-8 (1998)
15. A. J. McCoy, V. C. Epa and P. M. Colman: Electrostatic complementarity at protein/protein interfaces. *J Mol Biol* 268, 570-84 (1997)
16. S. A. Botti, C. E. Felder, J. L. Sussman and I. Silman: Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Eng* 11, 415-20 (1998)
17. G. Johnson and S. W. Moore: Human acetylcholinesterase binds to mouse laminin-1 and human collagen IV by an electrostatic mechanism at the peripheral anionic site. *Neurosci Lett* 337, 37-40 (2003)
18. R. R. Gabdoulline and R. C. Wade: Biomolecular diffusional association. *Curr Opin Struct Biol* 12, 204-13 (2002)
19. M. Stein, R. R. Gabdoulline and R. C. Wade, Integrating Structural and Kinetic Enzymatic Information

in Systems Biology, John von Neumann Institute for Computing, Juelich. (2006)

20. R. R. Gabdoulline and R. C. Wade: Simulation of the diffusional association of barnase and barstar. *Biophys J* 72, 1917-29 (1997)

21. R. R. Gabdoulline, R. C. Wade and D. Walther: MolSurfer: two-dimensional maps for navigating three-dimensional structures of proteins. *Trends Biochem Sci* 24, 285-7 (1999)

22. F. De Rienzo, R. R. Gabdoulline, M. C. Menziani, P. G. De Benedetti and R. C. Wade: Electrostatic analysis and Brownian dynamics simulation of the association of plastocyanin and cytochrome f. *Biophys J* 81, 3090-104 (2001)

23. E. D. Getzoff, D. E. Cabelli, C. L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci and R. A. Hallewell: Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature* 358, 347-51 (1992)

24. R. C. Wade, R. R. Gabdoulline, S. K. Luedemann and V. Lounnas: Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *Proc Natl Acad Sci U S A* 95, 5942-9 (1998)

25. K. Kinoshita and H. Nakamura: eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* 20, 1329-30 (2004)

26. P. J. Winn, J. N. D. Battey, K. Schleinkofer, A. Banerjee and R. C. Wade: Issues in high-throughput comparative modelling: a case study using the ubiquitin E2 conjugating enzymes. *Proteins* 58, 367-75 (2005)

27. X. Zhang, C. L. Bajaj, B. Kwon, T. J. Dolinsky, J. E. Nielsen and N. A. Baker: Application of New Multiresolution Methods for the Comparison of Biomolecular Electrostatic Properties in the Absence of Global Structural Similarity. *Multiscale Model Sim* 5, 1196-1213 (2006)

28. F. Fogolari, A. Brigo and H. Molinari: The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit* 15, 377-92 (2002)

29. R. C. Wade, R. R. Gabdoulline and F. De Rienzo: Protein Interaction Property Similarity Analysis. *International Journal of Quantum Chemistry* 83, 122-127 (2001)

30. N. Blomberg, R. R. Gabdoulline, M. Nilges and R. C. Wade: Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* 37, 379-87 (1999)

31. F. G. Whitby, G. Xia, C. M. Pickart and C. P. Hill: Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. *J Biol Chem* 273, 34983-91 (1998)

32. N. Saitou and M. Nei: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-25 (1987)

33. W. L. DeLano, The PyMOL Molecular Graphics System (2002) on World Wide Web <http://www.pymol.org>, (2002)

34. J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Lutty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Gagheri, L. R. Scott and J. A. McCammon: Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 91, 57-96 (1995)

35. J. Zhu and Z. Weng: FAST: a novel protein structure alignment algorithm. *Proteins* 58, 618-27 (2005)

36. J. M. Desterro, M. S. Rodriguez and R. T. Hay: SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. *Mol Cell* 2, 233-9 (1998)

37. Q. Liu, C. Jin, X. Liao, Z. Shen, D. J. Chen and Y. Chen: The binding interface between an E2 (UBC9) and a ubiquitin homologue (UBL1). *J Biol Chem* 274, 16979-87 (1999)

38. M. H. Tatham, S. Kim, B. Yu, E. Jaffray, J. Song, J. Zheng, M. S. Rodriguez, R. T. Hay and Y. Chen: Role of an N-terminal site of Ubc9 in SUMO-1, -2, and -3 binding and conjugation. *Biochemistry* 42, 9959-69 (2003)

39. K. P. Bencsath, M. S. Podgorski, V. R. Pagala, C. A. Slaughter and B. A. Schulman: Identification of a multifunctional binding site on Ubc9p required for Smt3p conjugation. *J Biol Chem* 277, 47938-45 (2002)

40. L. M. Lois and C. D. Lima: Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *Embo J* 24, 439-51 (2005)

41. D. T. Huang, A. Paydar, M. Zhuang, M. B. Waddell, J. M. Holton and B. A. Schulman: Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Mol Cell* 17, 341-50 (2005)

42. T. Miura, W. Klaus, B. Gsell, C. Miyamoto and H. Senn: Characterization of the binding interface between ubiquitin and class I human ubiquitin-conjugating enzyme 2b by multidimensional heteronuclear NMR spectroscopy in solution. *J Mol Biol* 290, 213-28 (1999)

43. K. S. Hamilton, M. J. Ellison, K. R. Barber, R. S. Williams, J. T. Huzil, S. McKenna, C. Ptak, M. Glover and G. S. Shaw: Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure (Camb)* 9, 897-904 (2001)

44. P. S. Brzovic, A. Lissounov, D. E. Christensen, D. W. Hoyt and R. E. Klevit: A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol Cell* 21, 873-80 (2006)

45. N. Merkley and G. S. Shaw: Solution structure of the flexible class II ubiquitin-conjugating enzyme Ubc1 provides insights for polyubiquitin chain assembly. *J Biol Chem* 279, 47139-47 (2004)

46. H. Walden, M. S. Podgorski and B. A. Schulman: Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* 422, 330-4 (2003)

47. K. R. Loeb and A. L. Haas: The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J Biol Chem* 267, 7806-13 (1992)

48. C. Zhao, S. L. Beaudenon, M. L. Kelley, M. B. Waddell, W. Yuan, B. A. Schulman, J. M. Huibregtse and R. M. Krug: The UbcH8 ubiquitin E2 enzyme is also the E2 enzyme for ISG15, an IFN-alpha/beta-induced ubiquitin-like protein. *Proc Natl Acad Sci U S A* 101, 7578-82 (2004)

49. W. Yuan and R. M. Krug: Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein. *Embo J* 20, 362-71 (2001)

50. E. Knight, Jr., D. Fahey, B. Cordova, M. Hillman, R. Kutny, N. Reich and D. Blomstrom: A 15-kDa interferon-induced protein is derived by COOH-terminal processing of a 17-kDa precursor. *J Biol Chem* 263, 4520-2 (1988)

51. L. Shen, C. Dong, H. Liu, J. H. Naismith and R. T. Hay: The structure of SENP1 SUMO-2 co-complex suggests a structural basis for discrimination between SUMO paralogues during processing. *Biochem J* (2006)
52. Z. Xu and S. W. Au: Mapping residues of SUMO precursors essential in fission yeast that is essential for the onset of anaphase in mitosis. *Mol Cell Biol* 17, 3388-97 (1997)
53. F. M. Townsley, A. Aristarkhov, S. Beck, A. Hershko and J. V. Ruderman: Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. *Proc Natl Acad Sci U S A* 94, 2362-7 (1997)
54. F. Osaka, H. Seino, T. Seno and F. Yamao: A ubiquitin-conjugating enzyme in fission yeast that is essential for the onset of anaphase in mitosis. *Mol Cell Biol* 17, 3388-97 (1997)
55. M. Tokumoto, Y. Nagahama and T. Tokumoto: Molecular cloning of cDNA encoding a cyclin-selective ubiquitin carrier protein (E2-C) from *Carassius auratus* (goldfish) and expression analysis of the cloned gene. *FEBS Lett* 458, 375-7 (1999)
56. H. Yu, R. W. King, J. M. Peters and M. W. Kirschner: Identification of a novel ubiquitin-conjugating enzyme involved in mitotic cyclin degradation. *Curr Biol* 6, 455-66 (1996)
57. R. W. King, R. J. Deshaies, J. M. Peters and M. W. Kirschner: How proteolysis drives the cell cycle. *Science* 274, 1652-9 (1996)
58. F. M. Townsley and J. V. Ruderman: Functional analysis of the *Saccharomyces cerevisiae* UBC11 gene. *Yeast* 14, 747-57 (1998)
59. P. Kaiser, S. Mandl, M. Schweiger and R. Schneider: Characterization of functionally independent domains in the human ubiquitin conjugating enzyme UbcH2. *FEBS Lett* 377, 193-6 (1995)
60. P. Kaiser, W. Seufert, L. Hofferer, B. Kofler, C. Sachsenmaier, H. Herzog, S. Jentsch, M. Schweiger and R. Schneider: A human ubiquitin-conjugating enzyme homologous to yeast UBC8. *J Biol Chem* 269, 8797-802 (1994)
61. S. Jentsch, J. P. McGrath and A. Varshavsky: The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* 329, 131-4 (1987)
62. M. G. Goebel, J. Yochim, S. Jentsch, J. P. McGrath, A. Varshavsky and B. Byers: The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. *Science* 241, 1331-5 (1988)
63. K. Robzyk, J. Recht and M. A. Osley: Rad6-dependent ubiquitination of histone H2B in yeast. *Science* 287, 501-4 (2000)
64. S. Qin, B. Nakajima, M. Nomura and S. M. Arfin: Cloning and characterization of a *Saccharomyces cerevisiae* gene encoding a new member of the ubiquitin-conjugating protein family. *J Biol Chem* 266, 15549-54 (1991)
65. P. Sung, S. Prakash and L. Prakash: The RAD6 protein of *Saccharomyces cerevisiae* polyubiquitinates histones, and its acidic domain mediates this activity. *Genes Dev* 2, 1476-85 (1988)
66. E. D. Lowe, N. Hasan, J. F. Trempe, L. Fonso, M. E. Noble, J. A. Endicott, L. N. Johnson and N. R. Brown: Structures of the Dsk2 UBL and UBA domains and their complex. *Acta Crystallogr D Biol Crystallogr* 62, 177-88 (2006)
67. C. N. Larsen, J. S. Price and K. D. Wilkinson: Substrate binding and catalysis by ubiquitin C-terminal hydrolases: identification of two active site residues. *Biochemistry* 35, 6735-44 (1996)
68. F. De Rienzo, R. R. Gabdoulline, M. C. Menziani and R. C. Wade: Blue copper proteins: a comparative analysis of their molecular interaction properties. *Protein Sci* 9, 1439-54 (2000)
69. M. Stein, R. R. Gabdoulline and R. C. Wade: Comparative cross-species analysis of biochemical pathways using protein interaction fields. *Personal communication* (2007)
70. I. Letunic, R. R. Copley, B. Pils, S. Pinkert, J. Schultz and P. Bork: SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 34, D257-60 (2006)
71. R. R. Gabdoulline, S. Ulbrich, S. Richter and R. C. Wade: ProSAT2--Protein Structure Annotation Server. *Nucleic Acids Res* 34, W79-83 (2006)
72. L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg: The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449-51 (2004)
73. R. W. Hooft, G. Vriend, C. Sander and E. E. Abola: Errors in protein structures. *Nature* 381, 272 (1996)

Abbreviations: SI_{Hodgkins} : Hodgkin similarity index; SI_{Carbo}: Carbo similarity index; UBL: ubiquitin like; UFP; ubiquitin family protein; MUB : membrane anchored ubiquitin fold protein; PIPSA : Protein Interaction Property Similarity Analysis

Key Words: Electrostatic Potential, Ubiquitin, Ubiquitin Like Protein, Ubiquitin Conjugating Enzyme, High Throughput Modelling, Structure-Based Design, Review

Send correspondence to: Dr Peter J Winn EML Research gGmbH, Villa Bosch, Schloss-Wolfsbrunnengasse 33, 69118 Heidelberg, Germany, Tel: +49 6221 533269, Fax: +49 6221 533298, E-mail: Peter.Winn@eml-r.villa-bosch.de

<http://www.bioscience.org/current/vol12.htm>