Genetica (2007) 129:45–55
DOI 10.1007/s10709-006-0032-3

REVIEW PAPER

Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions

Eric S. Haag

Received: 18 May 2004 / Accepted: 17 June 2005 / Published online: 16 November 2006

Abstract The evolution of molecules, developmental circuits, and new species are all characterized by the accumulation of incompatibilities between ancestors and descendants. When specific interactions between components are necessary at any of these levels, this requires compensatory coevolution. Theoretical treatments of compensatory evolution that only consider the endpoints predict that it should be rare because intermediate states are deleterious. However, empirical data suggest that compensatory evolution is common at all levels of molecular interaction. A general solution to this paradox is provided by plausible neutral or nearly neutral intermediates that possess informational redundancy. These intermediates provide an evolutionary path between coadapted allelic combinations. Although they allow incompatible end points to evolve, at no point was a deleterious mutation ever in need of compensation. As a result, what appears to be compensatory evolution may often actually be ''pseudocompensatory.'' Both theoretical and empirical studies indicate that pseudocompensation can speed the evolution of intergenic incompatibility, especially when driven by adaptation. However, under strong stabilizing selection the rate of pseudocompensatory evolution is still significant. Important examples of this process at work discussed here include the evolution of rRNA secondary structures, intra- and inter-protein interactions, and developmental genetic pathways. Future empirical work in this area should focus on comparing the details of intra- and intergenic interactions in closely related organisms.

Keywords Compensatory evolution · Developmental system drift · Epistasis · Genetic pathway evolution · Incompatibility · pseudocompensation · RNA structure

> ''Man is a theorizing animal. He is continually engaged in veiling the austerely beautiful outline of reality under myths and fancies of his own device. The truly scientific attitude, which no scientist can constantly preserve, is a passionate attachment to reality as such, whether it be bright or dark, mysterious or intelligible.''–JBS Haldane (1932, p170).

Introduction

The last few decades have revealed an important truth about evolution: it never stops. That this is so underlies the assumptions of many subfields of evolutionary biology: Without a reliable correlation, imperfect though it may be, between time and degree of divergence, phylogenetic systematics, evolutionary population genetics, and the emerging comparative genomics (both functional and informatic) would lose their central organizing principle. Kimura's assertion (e.g. 1983) that the bulk of sequence evolution is neutral or nearly so has been widely accepted as the reason for this crucial correlation at the molecular level. However, the relationship (or lack thereof) between this incessant change at the molecular level and the more irregular pace of morphological change has only recently become a tractable research topic. Writing in 1965, the

E. S. Haag (✉)
Department of Biology, University of Maryland,
Bldg. #144, College Park, MD 20742, USA
e-mail: ehaag@umd.edu

pioneering molecular evolutionist Emile Zuckerkandl enunciated two reasonable hypotheses:

> ...It is often said that evolution has been just as long for organisms that appear to have changed little as for those that have changed much; consequently it is held that the biochemistry of living animals is probably very different from that of their remote ancestors. My own view is that it is unlikely that selective forces would favor the stability of morphological characteristics without at the same time favoring the stability of biochemical characteristics, which are more fundamental.

Though Zuckerkandl favored the simpler idea of a direct correlation between molecular and morphological change, it now seems that the ''unlikely'' scenario is closer to the truth. Rather than being isolated from the constant change swirling around it, developmental regulatory pathways and their components often evolve significantly. This is true even in the face of morphological stasis, a phenomenon termed developmental system drift (DSD, see True and Haag 2001, for discussion). Though a major goal of evolutionary developmental biology is to explain the evolution of form, the adaptive fraction of molecular and developmental change may in fact be troublingly small (Haag and True 2001). This paper explores the idea that DSD is generally produced by compensatory evolution acting at different levels of biological organization.

Not only do molecular sequences change—so do the interactions between molecules and the organization of these interactions into higher order pathways. As with silent mutations in DNA, many of these changes may be neutral because they provide equally useful alternate solutions to a common need. But unlike silent mutations, neutrality is a synthetic feature of two loci. Compensatory changes would seem to require individual mutations that produce dysfunctional intermediates, classically envisioned by Sewall Wright as a fitness valley between adaptive peaks. Such intermediates should greatly reduce the power of stochastic processes to cause higher-order drift, and modeling has reinforced this impression (e.g. Kimura 1985; Barton 1989; Michalakis and Slatkin 1996; Phillips 1996; Stephan 1996).

One goal of this paper is to demonstrate that, despite the above theoretical expectation, compensatory change is rampant at all levels of molecular interaction, and thus is an important engine of incompatibility. Another is to suggest that the lack of congruence between theory and reality is due to an unrealistic assumption of the dominant models, the existence of only the endpoint states of the components of each interaction. Phillips (1996) has suggested that the only reasonable way to achieve compensatory evolution is through the relaxation of selection on intermediates, because ''mutations are too rare in small populations, and selection too effective in large populations.'' Reformulation of the problem in terms of specific molecular mechanisms often suggests intermediate states, in many cases transient, that would accomplish this relaxation of selection and eliminate the fitness cost of intermediate genotypes. These ''molecular stepping stones'' all have in common an aspect of excess capacity, an oft-noted facilitator of divergence (see discussion by Stoltzfus 1999). Where these intermediates are as fit as the incompatible endpoints then evolution is formally not compensatory, but if the intermediates are subsequently lost it may appear that it is. I dub this phenomenon ''pseudocompensation'' for clarity. This idea is related to Gavrilets' ''holey landscape'' view of adaptation (Gavrilets 1997), in that ''ridges'' of high fitness connect incompatible genotypes, but differs from it in that here the ridges are based upon transitional alleles of a single pair of interacting loci rather than the cumulative effect of many loci. I provide evidence that neutral intermediates may facilitate many of the apparently compensatory transitions in what Haldane (1931) called ''metastable systems,'' and identify specific research programs that will test their importance.

Compensatory evolution in the real world

The criteria for a potentially compensatory system are:

(1) Multiple potential sequence or structural solutions for each interacting set of elements, often inferred from the existence of highly polymorphic loci within a single species, or from highly divergent orthologous genes of closely related species.
(2) A significant cost of mismatched elements should be likely or experimentally demonstrable.

Below, examples of compensatory evolution at various levels are provided, and where possible likely intermediates between states are described.

Within a molecule

The most studied cases of intramolecular coevolution are in non-coding RNAs, such as ribosomal RNA, which nicely satisfy the criteria above. Most compensated variation in base-paired stems (intramolecular hairpins) is likely to be neutral, but non-complementary mismatches should be deleterious because they disrupt the helical structure. The small number of

available nucleotides makes possible realistic modeling of mutation and selection in populations (Muse 1995; Stephan 1996; Higgs 1998). Empirically, compensation in rRNA is recognized from the combination of secondary structure models and comparative sequence data (reviewed by Gutell et al. 2002). Experimental work with transgenic organisms has demonstrated that compensated mutations often function nearly as well the wild-type base pair, and better than mismatches. The ''instant evolution'' experiment of Morosyuk et al. (2000) offers an illuminating example (Fig 1).

In this study, pools of rRNA molecules with randomized nucleotides in a single stem-loop region, the ''690 loop,'' were screened for biological function in vivo (Fig 1A, B). In their system, growth depends upon detoxification of chloramphenicol (an antibiotic) in the growth medium by expression of the enzyme chloramphenicol acetyl-transferase (CAT). This in turn depends on the test rRNA with the randomized region, because the CAT gene encodes a modified mRNA with an atypical ribosome binding sequence that only the plasmid-encoded test rRNA can bind. This specificity is due to a minor change in a region distinct from the randomized stem. The test rRNA cannot translate other cellular mRNA due to the same modification, thus providing a tightly controlled assay for its function. The assay is quantified by determining the lowest concentration of chloramphenicol at which a plasmid can no longer support growth.

Figure 1B shows the minimum concentration of Cam required to stop growth of cells bearing plamids with the indicated base pair at positions 688/699. While presumably unpaired bases support some viability (e.g. A:C; gray bars), only the four canonical Watson–Crick base pairs and one particular non-canonical pair—G:U—were found in the most active subset (black bars). The unique capacity of the G:U pair seen in this assay has long been recognized, and indeed is present elsewhere in the natural sequence of the 690 loop (Fig 1A) and other rRNAs (Fig 1C). The partial function of completely mismatched bases is most likely explained by the simultaneous presence of other, second-site compensatory changes in the randomized part of these molecules. That the wild-type base pair often functions slightly better than alternative canonical base pairs in these systems implies that more subtle forms of compensation, such as those involving hydrophobic stacking interactions between adjacent base pairs, also play a role in allowing sequence change.

Theory suggests that direct compensation from one canonical base pair to another through a mispaired intermediate would be facilitated by tight linkage and the shielding from selection provided by the high copy number of rRNA genes (Kimura 1985; Higgs 1998). In addition, a relatively small number of mutational options are open to a base-paired ribonucleotide. Nevertheless, a plausible intermediate between Watson and Crick base pairs still exists. The study of Morosyuk et al. (2000) discussed above shows that the non-canonical G:U base pair is capable of substituting for canonical pairs with near-normal function in stems (Fig 1A, B). Comparative studies have shown very similar results, with the G:U pair being found as a common substitute for canonical pairs in closely related species (Fig 1C; Rousset et al. 1991; Hickson et al. 1996). Therefore, even in the best conditions imaginable for direct compensation, an intermediate state is nevertheless commonly used.

Intramolecular compensation in protein evolution will generally be more difficult to detect because of the diversity of amino acids and the lack of reliable ab initio structural predictions. However, progress has been made through both bioinformatic and experimental approaches. An example of the first is the application of the information theory concept of mutual information (MI) to detect significant levels of covariation in amino acids (Wollenberg and Atchley 2000). Where a three-dimensional structural model exists, sites with high degrees of MI are frequently packed near each other (Wollenberg and Atchley 2000; Atchley et al. 2000). A second, especially clever informatic approach has been to study cases where missense mutations that cause disease in humans are fixed in the wild-type sequence of other animals, known as ''compensated pathogenic deviations'' (CPDs; Kondrashov et al. 2002). CPDs are surprisingly common even within the mammals (Waterston et al. 2002), and have been estimated to comprise roughly 10% of deviations from the orthologous human sequence (Kondrashov et al. 2002). The inferred compensation is not necessarily intramolecular, but in several cases specific candidates can be found with the aid of a crystal structure (see Fig 2). MI and CPDs are thus useful tools to predict intra-protein interactions between amino acids that vary between homologs. The extent to which the imperfect covariation seen in interacting amino acids is due to alternate ''intermediates'' will require solution of structures of closely related homologs, which has not generally been a priority for crystallographers. Such studies will be motivated by interest in determining the variations in sequence–structure relationships that allow protein divergence to occur.

Experimental work is uncovering compensatory changes in proteins over both laboratory and phylogenetic time scales. Over 40 years ago, Crick et al. (1961)

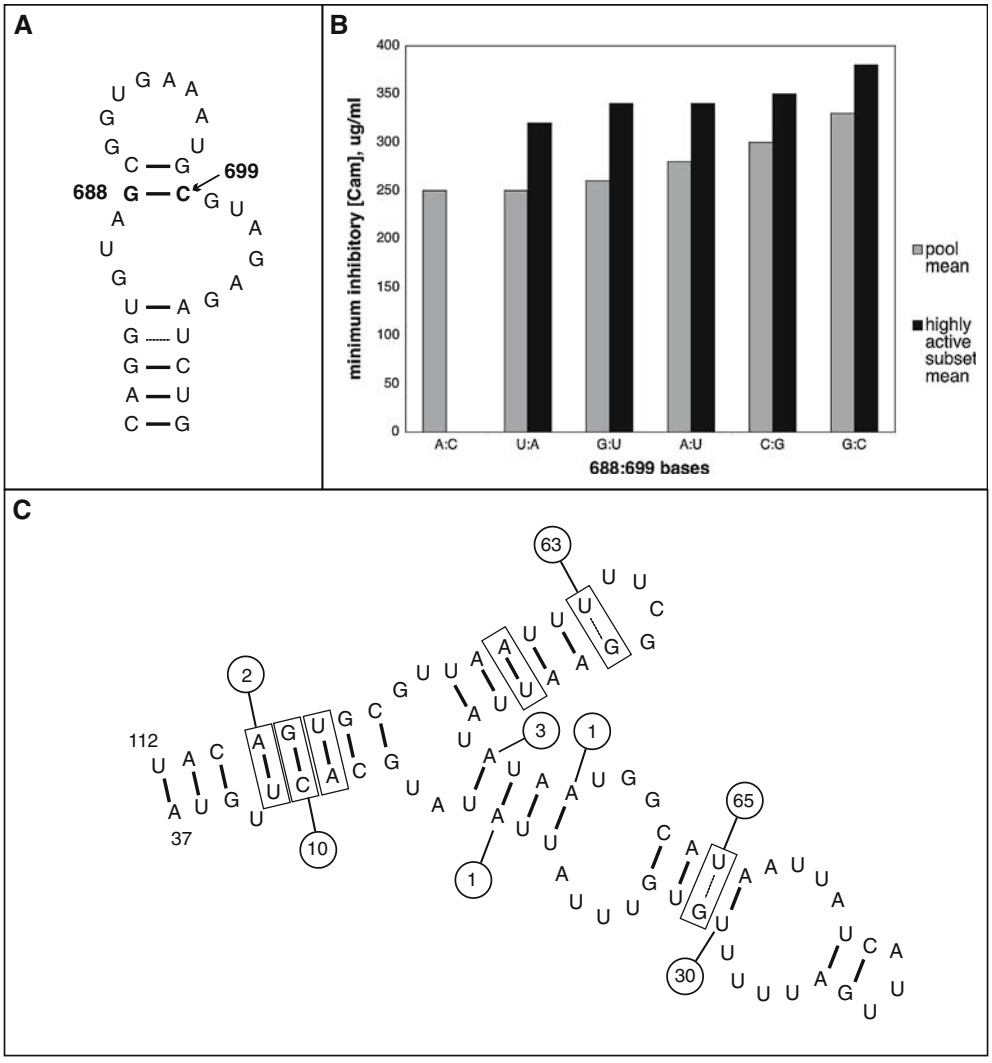

Fig. 1 Compensatory changes in RNA secondary structure mediated by G:U intermediates. (**A**) The 690 hairpin of the *Escherichia coli* 16S RNA. The 688/699 base pair appears in bold. (**B**) Data from the ''instant evolution'' experiment of Morosyuk et al. (2000), in which pools of molecules with randomized nucleotides in a normally base-paired region were screened for biological function by their ability to support growth in the presence of chloramphenicol. The assay relies on the plasmid-encoded rRNA being alone capable of translating a mutant chloramphenicol acetyl-transferase (Cam) mRNA bearing an atypical ribosome binding sequence. The chart shows the minimum concentration of Cam required to stop growth of cells bearing clones with the indicated base pair at positions 688/699. While unpaired bases support some viability (e.g. A:C; gray bars), only canonical and the G:U base pairs were found in the most active subset (black bars). The partial function of mismatched bases is most likely explained by the simultaneous presence of other, second-site compensatory changes in these molecules. (**C**) A portion of the secondary structure for the D2 region of the large subunit rRNA of *Drosophila* studied by Rousset et al. (1991), in which they examined 82 drosophilid species. Rectangles indicate base pairs exhibiting compensatory changes between different canonical Watson–Crick pairs. Circled numbers point to the number of taxa in which a G:U base pair is seen at a given position. The two uncircled numbers at the left are nucleotide positions. Note that in several cases the G:U pair is apparently a transient state, while in others it is the most common

used intragenic suppressors of mutations in phage T4 to discover the triplet basis for genetic code. Although these suppressors were small indels that corrected reading frame shifts, since then examples of intragenic suppressors of missense mutations have been found, for example in the T4 lysozyme and in *E. coli* enzymes (e.g. Poteete et al. 1991; Chen et al. 1996). These studies were based on experimentally induced mutations, but comparative methods are also proving informative. For example, Zhang and Rosenberg (2002) examined the recently duplicated eosinphil-derived neurotoxin (EDN) ribonuclease of humans and old world monkeys. By combining phylogeny-based ancestral sequence reconstructions with ribonuclease activity assays, they showed that two derived EDN residues function synergistically to boost the ribonuclease

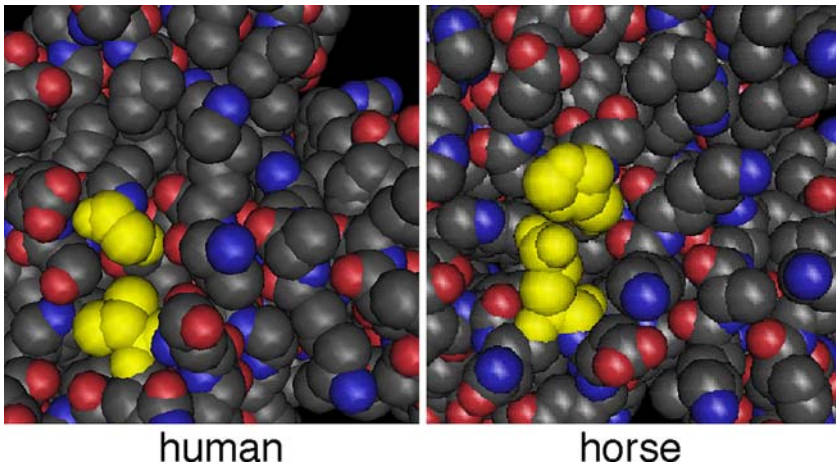

Fig. 2 A likely compensated pathogenic deviation at positions 20 and 69 of β-hemoglobin in humans and horses. The wild-type human amino acids are Val20, Gly69, and the mutation Val20Glu is pathogenic. Glu20 is the wild-type amino acid in the horse, presumably because it is accompanied by a compensatory change of residue 69 to histidine (Kondrashov et al. 2002). In the image, the two residues in question are in yellow, with position 69 towards the top. In surrounding residues, black represents carbon atoms, blue nitrogen, and red oxygen

activity of the inferred ancestor by over an order of magnitude. The side chains of these residues interact in the EDN tertiary structure, which suggests that they represent an example of adaptive, compensatory, intramolecular evolution.

Another, more explicitly developmentally relevant context in which intramolecular compensatory evolution is important is in *cis*-regulatory DNA sequences that govern temporal and spatial patterns of gene expression. The details of regulatory DNA evolution have recently been reviewed by Wray et al. (2003), and the accompanying paper by Hahn further explores this subject in detail. Here I will focus on how compensatory change might work in a promoter. Compared to RNA secondary structure the situation is more fluid, because compensation is not localized to a particular nucleotide. In principle, this could greatly speed sequence evolution by offering many more routes to successful compensation, and in practice it appears that this is the case. Even closely related species can have essentially unalignable *cis*-regulatory DNA that nevertheless drives conserved expression patterns (Tamarina et al. 1997; Ludwig et al. 1998; Romano and Wray 2003). This effect is especially pronounced in organisms with large population sizes, which is predicted to facilitate the evolution of adaptive compensated genotypes (Carter and Wagner 2002).

The work of Ludwig et al. (2000) with the *Drosophila evenskipped* stripe 2 enhancer showed that compensation acts over distances large enough to be disrupted through the production of chimaeric promoter constructs assembled from divergent, but homologous, sequences. Thus, functionally equivalent modules seem to ''move'' over time. A simple mechanism for this apparent movement is via an intermediate state in which redundant transcription factor binding sites accumulate and are subsequently lost differentially between lineages. This implies that population-level variation in enhancers should exist, and a recent study of Rockman and Wray (2002) has identified it in a large set of human genes. Although these variants are circulating in viable people, in many cases they are associated with significant differences in transcription levels. Enhancer polymorphisms that alter a gene's ancestral expression level may be mildly deleterious or truly neutral, and which is true may frequently depend on genetic context. In either case, they could represent as-yet-uncompensated first steps towards fixed divergence in enhancer architecture. Comparisons with the promoters of chimpanzee orthologues may provide a valuable perspective on these polymorphisms.

Between products encoded by linked genes

Much of the business of cells and tissues requires the specific binding of one gene product by another. Compensatory change in binding partners has been encountered here as well, most notably in reproduction-related proteins that are diverging at rates much higher than the norm. These represent the most extreme examples of intermolecular compensatory evolution, and thus hold unique promise as systems in which to analyze how interacting molecules can maintain their essential interactions with a constantly changing sequence. It should be noted that how this occurs is a fundamental problem that will greatly impact structural biology per se.

Theoretical studies agree that tight linkage greatly facilitates compensatory evolution (e.g. Kimura 1985; Phillips 1996). It may not be surprising, then, that one of the most spectacular intermolecular examples occurs in the two tightly linked genes of the *Brassica* self-incompatibility (SI) locus, termed *S* (Kachroo et al. 2002). Each *S* haplotype encodes one receptor serine/threonine kinase (*SRK*), expressed in the stigma, and one cysteine-rich ligand (*SCR*) for *SRK*, which is expressed only in anthers. Each of the more than 50 haplotypes encode divergent receptor-ligand pairs that interact with tight specificity (Watanabe et al. 2000; Kachroo et al. 2001).

In a two-component SI system, linkage is clearly crucial to prevent selfing or inappropriate rejection of unrelated pollen, and this constraint must contribute to the rapid compensatory evolution of haplotypes. However, the great allelic diversity also suggests that new alleles are produced frequently from old ones. One proposal for how this might work in the solonaceae (Matton et al. 1999) is a stepping stone-type model, in which the stigma component first evolves dual specificity for both its ancestral partner allele and a potential new derivative. A subsequent mutation in the pollen component could then take advantage of this excess capacity, after which subsequent loss of the ancestral compatibility of the stigma component would complete the generation of a novel haplotype.

The simple model of Matton et al. (1999) preserves functional SI throughout the generation of new alleles, which would seem a virtue. However, it was criticized by Charlesworth (2000) because it required a potentially slow succession of mutations, and by Uyenoyama and Newbigin (2000) because deterministic modeling suggested that nonreciprocal rejection of intermediate genotypes would tend to block the increase in haplotype diversity. More recently, Uyenoyama et al. (2001) proposed a different scenario, in which the first step is a mutation disabling pollen recognition entirely, followed by reestablishment of SI with altered specificity. However, this route also tends to replace an old allele with a new one unless population subdivision shelters the former. Resolution of this interesting question will require further experimental work to determine which model's assumptions are closest to reality. For example, it may now be possible to measure the extent of inbreeding depression in wild isolates of self-incompatible model plants through the direct disabling of SI with mutations or transgenes.

Between products of unlinked genes

Proteins encoded by unlinked genes can also undergo rapid compensatory evolution. Now-classic examples include interacting fertilization-related proteins expressed in the sperm and eggs of marine invertebrates, which have recently been reviewed by Swanson and Vacquier (2002). In these cases there is evidence that positive selection is driving the sequence change (e.g. Swanson and Vacquier 1995; Hellberg et al. 2000), presumably due to either sexual selection within species or reinforcement of isolation between sympatric species. The end result is a system that seems to be engaged in pointless divergence from a biochemical point of view, but which is in fact at the front lines of adaptation.

Another class of unlinked genes whose products show rapid compensatory evolution has emerged in the signal transduction pathway mediating sex determination in nematodes. The membrane receptor TRA-2 promotes female cell fates in all species of *Caenorhabditis* surveyed thus far (Hodgkin and Brenner 1977; Kuwabara 1996; Haag and Kimble 2000). Crucial to its function are physical interactions with another female-promoting protein, the transcription factor TRA-1 (Wang and Kimble 2001; Lum et al. 2000), and the male-promoting cytoplasmic protein FEM-3 (Mehra et al. 1999). When the homologues of TRA-2 from *C. elegans, C. briggsae*, and *C. remanei* were aligned (Haag and Kimble 2000), the region defined as the minimal FEM-3-binding domain coincided almost perfectly with a region of marked hypervariability. This included substantial length differences and frequent substitutions.

The spotlight thus turned to FEM-3, which had eluded cloning by hybridization-based approaches in non-*elegans* species because of poor sequence conservation. The *C. elegans* genome sequence opened a synteny-based strategy (Kuwabara and Shah 1994), which was used to clone *fem-3* homologs from *C. briggsae* and *C. remanei* (Haag et al. 2002). Alignment of FEM-3 homologs showed that approximately 2/3 of its amino acids differed in each pairwise comparison, with frequent indels as well. Thus, FEM-3 behaves similarly to the hypervariable portion of TRA-2c along its entire length. Despite this divergence, all three TRA-2c/FEM-3 pairs interact strongly within a species. Importantly, though, no cross-species interaction is seen, which demonstrates that compensatory evolution, and not a complete lack of constraint, is at work. A similar species-specific conservation of interaction has also been found for the TRA-2c/TRA-1 interaction in *C. briggsae* and *C. elegans* (Wang and Kimble 2001).

Given the importance of these protein–protein interactions to development and reproduction, how and why are they evolving so quickly? Two broad explanations have been invoked: positive selection and relaxed constraint (Stothard and Pilgrim 2003).

The case of fertilization proteins discussed above seems to be a case of the former, while for sex determination it is still unclear. However, the peculiar epistasis characterizing the population-level evolution of interacting gene products must also be addressed in either case. Haag and Molla (2005) have used population simulations to model the allele substitutions that occur during the evolution of incompatibilities via compensation. Much like the SI model of described above (Matton et al. 1999), their model invokes a ''no-cost half-site'' scenario, in which polymorphisms that could eventually participate in a novel interaction are essentially neutral (Fig 3). Evolution of second, matching site might then make any one of the ancestral bonds expendable, and soon an incompatibility between ancestor and descendant may result. Such intermediate-facilitated compensation can occur in both neutral and adaptive forms, but adaptation (also called ''supercompensation,'' Phillips et al. (2000)) speeds evolution in all cases. Interestingly, the simulations show that the intermediate alleles actually slow down the adaptive evolution of incompatibility (relative to the no-intermediate, two-allele case) if the increased fitness of the derived, compensated genotype is derived solely from gain of a new bond. This is because the intermediates themselves have high fitness in this case.

The above model predicts that neutral or nearly neutral polymorphisms must exist at or near the interface between interacting proteins. In addition, while the posited initial, neutral mutation facilitating compensation may be polymorphic for long periods, when the compensatory change arises both will rapidly reach fixation. To examine whether such polymorphisms exist, Haag and Ackerman (2005) examined *fem-3* in 11 wild isolates of the outcrossing species *C. remanei*, which is known to harbor more intraspecies variation than its selfing congeners (Graustein et al. 2002). They found that *Cr*-FEM-3 harbors an unusually large amount of amino acid polymorphism along its entire length, mirroring (albeit to a lesser degree) its variation between species. It is reasonable to suppose that at least some of the polymorphic residues are at or near the dimerization interface, but structural characterization

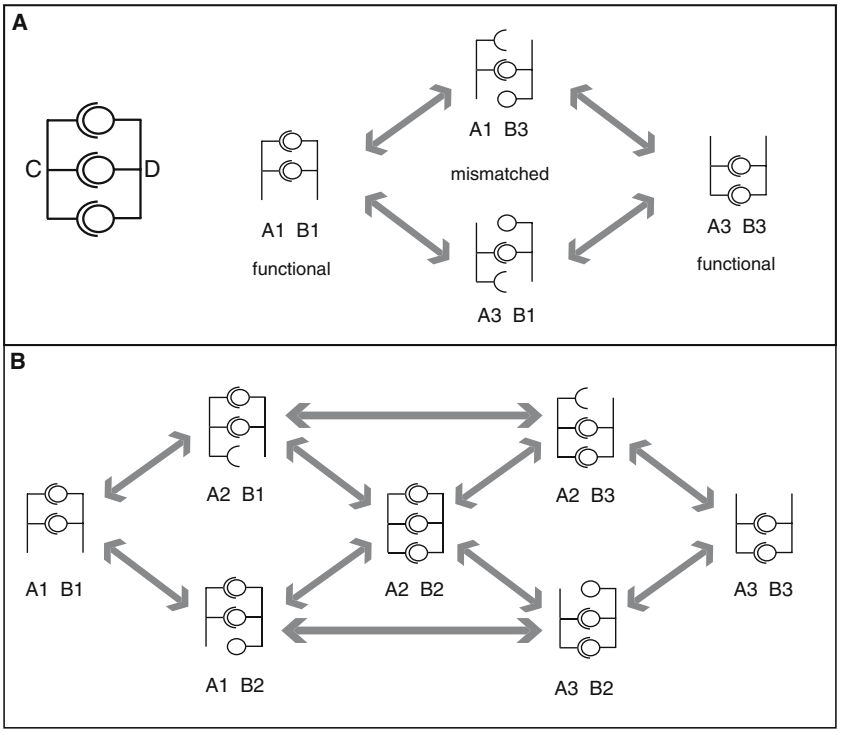

Fig. 3 Models for compensatory evolution in interacting gene products. (**A**) The traditional model, in which only alleles of the ancestral and compensated descendant states are included. At left is a cartoon of two interacting gene products, A and B, which have the potential to interact productively using any two of the three possible bonds. Considering only the A1B1 and A3B3 endpoint alleles predicts the need for either mismatched intermediates or simultaneous double mutation. (**B**) The neutral intermediates model analyzed by Haag and Molla (2005). In this model, an intermediate allele type, A2/B2, is posited that can interact productively with both ancestral and descendant alleles. Use of such intermediates requires more mutational steps to move from the A1B1 to A3B3, but in most cases the avoidance of mismatched genotypes more than makes up for it

of the FEM-3/TRA-2c complex is clearly needed clarify the situation. No evidence of local adaptation or extensive linkage disequilibrium was found in geographically distant isolates, so lack of constraint may explain the bulk of FEM-3 sequence divergence between taxa. Sites constrained by their interaction with TRA-2c may therefore be crucial, but yet relatively few in number.

Although the cases discussed above show that protein–protein interfaces can readily evolve, and in some cases do so more rapidly than neighboring non-interacting domains, it still presumably requires very particular mutations. It would therefore seem that in general the evolution of molecules participating in many interactions would be slowed, at least at the key interacting sites. Fraser et al. (2002) have presented evidence of exactly this in a comparison of yeast and *C. elegans* proteins. However, the magnitude of the effect is surprisingly small, and based on an extremely divergent set of homologs. A more recent analysis of a *Saccharomyces–Schizosaccharomyces* comparison (both yeasts) found a similarly weak correlation in all but the most extremely interactive proteins (Jordan et al. 2003). This can be explained in two ways, which are not mutually exclusive. First, compensatory changes in interacting proteins may not be much slower than other sorts of changes. For example, sites required for intra-molecular structure are similarly constrained. Alternatively, sites mediating inter-molecular interactions may indeed be relatively slow to evolve, but these sites are a tiny fraction of total sites.

Informational compensation: the modification of developmental pathways

The importance of molecular-level interaction in developmental biology has been nicely summarized by Johnson and Porter (2000):

> The overriding principle arising from the study of gene regulation during development is that gene interaction is of essence: the phenotype is created not only from the structural properties of the individual genes but also in very large part by the interactions among them.

There is ample evidence that particular regulatory connections, and even whole genes, can come in or out of the networks of interacting genes that form developmental pathways (reviewed by True and Haag 2001; Wilkins 2002). Similar to the intermolecular interactions, there cannot be any possibility for dysfunction at any point along a species' history. It therefore seems reasonable to again posit the establishment of redundancy as a necessary first step, followed by differential loss in isolated lineages. In general form this can be summarized as:

$$A \rightarrow A + B \rightarrow B$$

Such redundancy could come from two broad sources. One is the duplication of genes, in which case A and B above would be paralogs. The differential partitioning of functions between paralogs or differential loss of copies can facilitate divergence between taxa without loss of function at any point (Force et al. 1999; Lynch and Force 2000). Given the consistent function of homologous genes in the pathway, this latter process may often go unrecognized, appearing instead to be a "simple" movement of genes from syntenic to novel locations in the genome (Lynch et al. 2001).

Another source of redundancy can come from parallel pathways, which in turn may have been generated by a cooption of a gene or set of genes into a new function. In this case, A and B represent distinct, non-homologous gene networks that, perhaps through the regulation of common target genes, come to promote the same developmental fate. Evidence of such redundancy comes from enhancer or synthetic lethal screens. While such screens are commonplace in yeast and bacteria, they can also find unexpected new developmental functions for well-studied pathways in animals (e.g. Fay et al. 2003).

Recently, Johnson and Porter (2000; this volume) have modeled how networks of genes that regulate the quantitative output of a target gene evolve under directional selection. Their models include multiple, largely redundant regulators that act in series, as with transcription factors that regulate other transcription factors. The mathematical product of the gene products' activities determines the "phenotype." Their simulations demonstrate how isolated populations with identical starting genotypes can respond to directional selection in different ways, eventually producing hybrid incompatibilities. The key to this process is that the initial redundancy can be shed in any number of ways that are unconstrained even under selection. Although there is especially rapid divergence in pathways under directional selection, when a pleiotropic regulator affects both a directionally selected trait and another under stabilizing selection, the genetic control of the latter often diverges as well (Johnson and Porter 2005). This result therefore provides the first theoretical basis for the consistent evolution of DSD in gene regulatory networks, and suggests that the role of pleiotropy should feature prominently in future empirical studies.

Conclusions

An attempt has been made in this paper to make the point that, at the molecular level anyway, we may not need to wring our hands over the difficulty of compensatory mutation. The availability of low-cost molecular stepping stones may allow relatively easy, pseudocompensatory divergence of interacting genetic elements, even without tight linkage or strong selection. But with linkage and/or selection the process can go even faster, and this may explain why some of the most impressive cases of compensatory evolution involve linked and/or reproduction-related loci. If pseudocompensatory change is so widespread, then documenting the molecular and biochemical details of how it works will be of great interest to all biologists. Two related priorities for future work include:

Evolutionary structural biology

The complementary information provided by structure and variation has been used to great effect in the RNA field. Less has been done with the more challenging cases of proteins. Within species and between closely related species, some progress has been made through mapping variable residues onto a single representative structure. In some of the most interesting and extreme cases, however, it will be necessary to determine structures independently for each taxon, as has been done for abalone sperm lysins (Kresge et al. 2001). This will provide an exciting way to study the rules of protein higher-order structure, as well as give us some important insight into how compensation and pseudocompensation occur.

Comparative developmental genetics in sister species

We need to untangle the details of developmental genetic pathways in closely related species. Only through such an approach will we discover the ''quanta'' of change that compose larger leaps in developmental evolution. A large number of sister species pairs are being investigated, but often primarily through studies of gene expression or with transient assays for gene function, such as RNA interference. A more complete view of how development evolves will require the development of satellite models that have *bona fide* genetic tools that can match the rigor of their more famous relatives. This work has begun in earnest for a number of sister species of genetic model organisms, such as *Xenopus tropicalis* (Hirsch et al. 2002), *Caenorhabditis briggsae* (e.g. LaMunyon and Ward 1997; Hill et al. 2006), the more distant *C. elegans* relative *Pristionchus pacificus* (Srinivasan et al. 2003), *Arabidopsis lyrata* (Schierup et al. 2001), and *Drosophila simulans* (Sturtevant 1929; Barker and Moth 2001).

Acknowledgements The author thanks Arlin Stoltzfus for excellent conversations, and Allan Force and Norman Johnson for sharing relevant unpublished manuscripts. This work was supported from startup funds from the University of Maryland and an RE Powe, Jr. Faculty Enhancement Award from the Oak Ridge Associated Universities.

References

Atchley W, Wollenberg K, Fitch W, Terhalle W, Dress A (2000). Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. Mol Biol Evol 17:164–178

Barker J, Moth J (2001). Linkage maps of *D. simulans*: an update of Sturtevant (1929) with additional loci. Dros Inf Serv 84:205–206

Barton N. (1989). The divergence of a polygenic system subject to stabilizing selection, mutation, and drift. Gent Res Camb 54:59–77

Carter A, Wagner G (2002). Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc R Soc Lond B 269:953–960

Charlesworth D. (2000). How can two-gene models of self-incompatibility generate new specificities? Plant Cell 12:309–310

Chen R, Grobler J, Hurley J, Dean A (1996). Second-site suppression of regulatory phosphorylation in *Escherichia coli* isocitrate dehydrogenase. Protein Sci 5:287–295

Crick F, Barnett L, Brenner S, Watts-Tobin R. (1961). General nature of the genetic code for proteins. Nature 192:1227–1232

Fay D, Large E, Han M, Darland M (2003). *lin-35*/Rb and *ubc-18*, an E2 ubiquitin-conjugating enzyme, function redundantly to control pharyngeal morphogenesis in *C. elegans*. Development 130:3319–3330

Force A, Lynch M, Pickett F, Amores A, Yan Y, Postlethwait J (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Fraser H, Hirsh A, Steinmetz L, Scharfe C, Feldman M (2002). Evolutionary rate in the protein interaction network. Science 296:750–752

Gavrilets S. (1997). Evolution and speciation on holey adaptive landscapes. Trends Ecol Evol 12:307–312

Graustein A, Gaspar JM, Walters JR, Palopoli MF (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics 161:99–107

Gutell R, Lee J, Cannone J (2002). The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol 12:301–310

Haag E, Ackerman A (2005) Intraspecific variation in *fem-3*and *tra-2*, two rapidly coevolving nematode sex-determining genes. *Gene* pp TBA

Haag E, Kimble J (2000). Regulatory elements required for development of *C. elegans* hermaphrodites are conserved in the *tra*-2 homologue of C. *remanei*, a male/female sister species. Genetics 155:105–116

Haag E, Molla M (2005) Compensatory evolution of interacting gene products through multifunctional intermediates. Evolution 59:1620–1632

Haag E, True J (2001). From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. Evolution 55:1077–1084

Haag E, Wang S, Kimble J (2002). Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. Curr Biol 12:2035–2041

Haldane J. (1931). A mathematical theory of natural selection. Part VIII: metastable populations. Proc Camb Phil Soc 27:137–143

Haldane J (1932). The Causes of Evolution. Cornell Univ. Press, Ithaca, NY

Hellberg M, Moy G, Vacquier V (2000). Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. Mol Biol Evol 17:458–466

Hickson R, Simon C, Cooper A, Spicer G, Sullivan J, Penny D (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol Biol Evol 13:150–169

Higgs P (1998). Compensatory neutral mutations and the evolution of RNA. Genetica 102/103:91–101

Hill R, Carvalho C, Salogiannis J, Schlager B, Pilgrim D, Haag E (2006) Genetic flexibility in the convergent evolution of hermaphroditism in Caenorhabditis nematodes. Dev Cell 10:531–538

Hirsch N, Zimmerman L, Grainger R (2002). *Xenopus*, the next generation: *X. tropicalis* genetics and genomics. Dev Dyn 225:422–433

Hodgkin J, Brenner S (1977). Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. Genetics 86:275–287

Johnson N, Porter A (2000). Speciation via parallel, directional selection on regulatory genetic pathways. J Theor Biol 205:527–542

Johnson NA, Porter AH (2006) Evolution of branched regulatory genetic pathways: directional selection on pleiotropic loci accelerates developmental system drift. Genetica (in press)

Jordan I, Wolf Y, Koonin E (2003). No simple dependence between protein evolution rate and the number of protein–protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol 3:1–8

Kachroo A, Nasrallah M, Nasrallah J (2002). Self-incompatibility in the Brassicaceae: receptor-ligand signaling and cell–cell communication. Plant Cell 14(Suppl): S227–S238

Kachroo A, Schopfer C, Nasrallah M, Nasrallah J (2001). Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293:1824–1826

Kimura M (1983) The neutral theory of molecular evolution, Cambridge University Press

Kimura M (1985). The role of compensatory neutral mutations in molecular evolution. J Genet 64:7–19

Kondrashov A, Sunyaev S, Kondrashov F (2002). Dobzhansky-Muller incompatibilities in protein evolution. Proc Natl Acad Sci USA 99:14878–14883

Kresge N, Vacquier V, Stout D (2001). Abalone lysin: the dissolving and evolving sperm protein. BioEssays 23:95–103

Kuwabara PE (1996). Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. Genetics 144:597–607

Kuwabara PE, Shah S (1994). Cloning by synteny: identifying *C. briggsae* homologues of C. *elegans* genes. Nucleic Acids Res 22:4414–4418

LaMunyon C, Ward S (1997). Increased competitiveness of nematode sperm bearing the male X chromosome. Proc Natl Acad Sci USA 94:185–189

Ludwig M, Bergman C, Patel N, Kreitman M (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403:564–567

Ludwig M, Patel N, Kreitman M (1998). Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development 125:949–958

Lum D, Kuwabara P, Zarkower D, Spence A (2000). Direct protein–protein interaction between the intracellular domain of TRA-2 and the transcription factor TRA-1A modulates feminizing activity in *C. elegans*. Genes Dev 14:3153–3165

Lynch M, Force A (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473

Lynch M, O'Hely M, Walsh B, Force A (2001). The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804

Matton D, Luu D, Xike Q, Bertrand C, Morse D, Cappadocia M (1999). The production of an S-RNase with dual specificity suggests a novel hypothesis for the generation of new S-alleles. Plant Cell 11:2087–2097

Mehra A, Gaudet J, Heck L, Kuwabara PE, Spence AM (1999). Negative regulation of male development in *Caenorhabditis elegans* by a protein–protein interaction between TRA-2A and FEM-3. Genes Dev 13:1453–1463

Michalakis Y, Slatkin M (1996). Interaction of selection and recombination in the fixation of negative-epistatic genes. Genet Res Camb 67:257–269

Morosyuk S, Lee K, Santalucia JJ, Cunningham P (2000). Structure and function of the conserved 690 hairpin in *Escherichia coli* 16S ribosomal RNA: analysis of the stem nucleotides. J Mol Biol 300:113–126

Muse S (1995). Evolutionary analyses of DNA sequences subject to constraints of secondary structure. Genetics 139:1429–1439

Phillips P (1996). Waiting for a compensatory mutation: phase zero of the shifting-balance process. Genet Res Camb 67:271–283

Phillips P, Otto S, Whitlock M (2000) Beyond the Average. In: Wolf J, Brodie III ED, Wade MJ (eds) Epistasis and the Evolutionary Process. Oxford University Press, Oxford, pp20–38

Poteete A, Dao-Pin S, Nicholson H, Matthews B (1991). Second-site revertants of an inactive T4 lysozyme mutant restore activity by restructuring the active site cleft. Biochemistry 30:1425–1432

Rockman M, Wray G (2002). Abundant raw material for *cis*-regulatory evolution in humans. Mol Biol Evol 19:1991–2004

Romano L, Wray G (2003). Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both *cis* and *trans*-acting componenets of transcriptional regulation. Development 130:4187–4199

Rousset F, Pelandakis M, Solignac M (1991). Evolution of compensatory substitutions through G:U intermediate stat in *Drosophila* rRNA. Proc Natl Acad Sci USA 88:10032–10036

Schierup M, Mable B, Awadalla P, Charlesworth D (2001) Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. Genetics 158:387–399

Srinivasan J, Sinz W, Jesse T, Wiggers-Perebolte L, Jansen K, Buntjer J, van der Meulen M, Sommer R (2003) An integrated physical and genetic map of the nematode *Pristionchus pacificus*. Mol Genet Genomics 269:715–722

Stephan W (1996) The rate of compensatory evolution. Genetics 144:419–426

Stoltzfus A (1999) On the possibility of constructive neutral evolution. J Mol Evol 49:169–181

Stothard P, Pilgrim D (2003) Sex determination gene and pathway evolution in nematodes. BioEssays 25:221–231

Sturtevant A (1929) The genetics of *Drosophila simulans*. Carnegie Inst Wash Publ 339:50–61

Swanson W, Vacquier V (1995) Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc Natl Acad Sci USA 92:4957–4961

Swanson W, Vacquier V (2002) The rapid evolution of reproductive proteins. Nature Rev Gen 3:137–144

Tamarina N, Ludwig M, Richmond R (1997) Divergent and conserved features in the spatial expression of the *Drosophila pseudoobscura* esterase-5B gene and the esterase-6 gene of *Drosophila melanogaster*. Proc Natl Acad Sci USA 94:7735–7741

True J, Haag E (2001) Developmental system drift and flexibility in evolutionary trajectories. Evol Dev 3:109–119

Uyenoyama M, Newbigin E (2000) Evolutionary dynamics of dual-specificity self-incompatibility alleles. Plant Cell 12:310–312

Uyenoyama M, Zhang Y, Newbigin E (2001) On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. Genetics 157:1805–1817

Wang S, Kimble J (2001) The TRA-1 transcription factor binds TRA-2 to regulate sexual fates in *Caenorhabditis elegans*. EMBO J 20:1363–1372

Watanabe M, Ito A, Takada Y, Ninomiya C, Kakizakia T, Takahata Y, Hatakeyama K, Hinata K, Suzuki G, Takasaki T, Satta Y, Shiba H, Takayama S, Isogai A (2000) Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestri*s (syn. *rapa*) L. FEBS Lett 473:139–144

Waterston R, Lindblad-Toh K, Birney E, a. m. o. o. t. M.G.S. Consortium. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Wilkins A (2002) The Evolution of Developmental Pathways. Sinauer, Sunderland, MA

Wollenberg K, Atchley W (2000) Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. Proc Natl Acad Sci USA 97:3288–3291

Wray G, Hahn M, Abouheif E, Balhoff J, Pizer M, Rockman M, Romano L (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20:1377–1419

Zhang J, Rosenberg H (2002) Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc Natl Acad Sci USA 99:5486–5491

Zuckerkandl E (1965) The evolution of hemoglobin. Sci Am. 212:110–118