

Searching for Molecular Solutions – Cited Notes

CHAPTER 3

These Files contain details on all references to this ftp site within **Chapter 3** of *Searching for Molecular Solutions*. The page numbers of the book where the reference is made are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.

Contents:

Book Reference Page Number	Page Number in this File	Section	
		No.	Title
76	2	<u>3</u>	MHC-peptide Binding
84	4	<u>4</u>	DNA-binding protein design
92	23	<u>5</u>	Neural Diversity

Section 3: **MHC – Peptide Binding**

Cited on p. 76 of *Searching for Molecular Solutions*

(in Figure legend of Fig. 3.4; MHC-peptide binding structures)

This section extends the description of MHC-peptide recognition by the T cell receptor in *Searching for Molecular Solutions*, by comparing the recognition of MHC Class I peptide complexes with those of Class II (Fig. 3.Na).

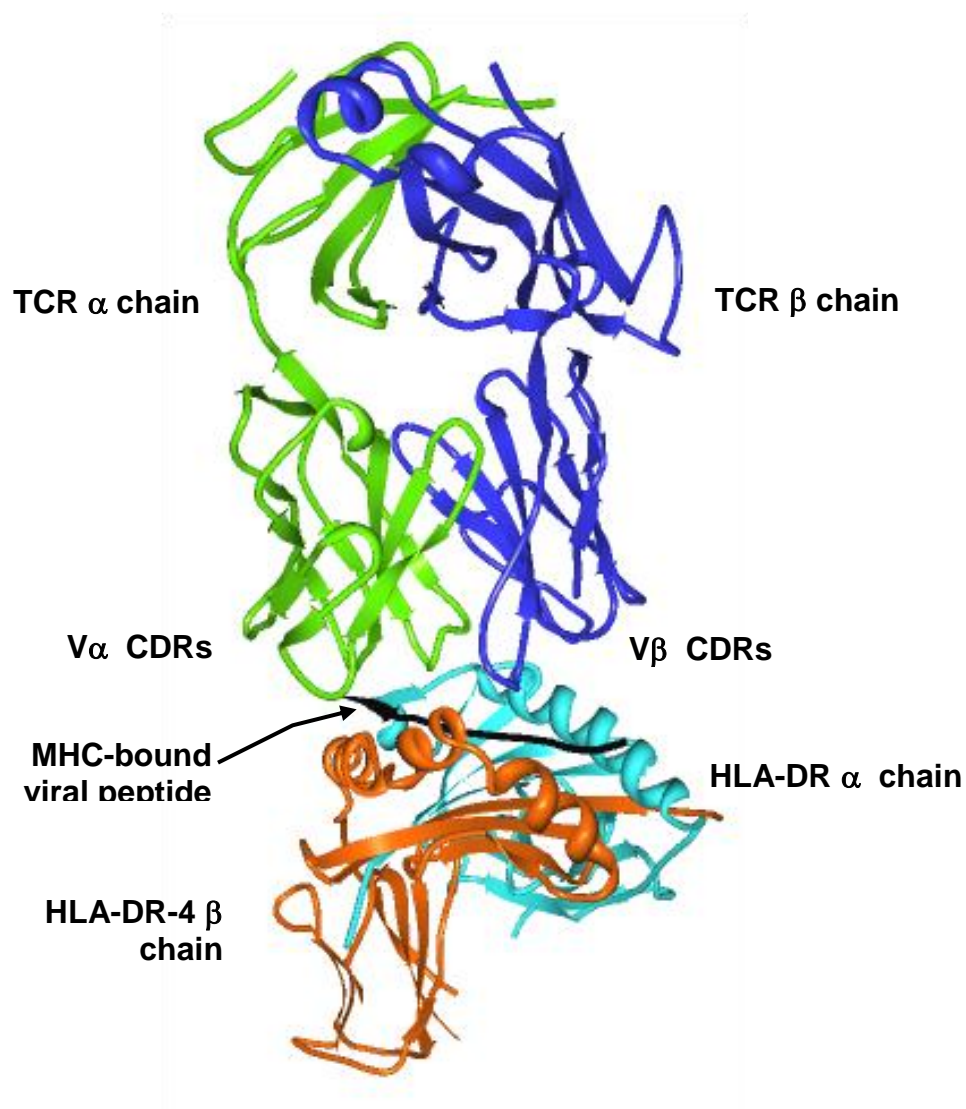


Fig. 3.Na

Structure of a complex between MHC Class II (human HLA-DR α chain [light blue] and HLA-DR-4 β chain [orange]) binding a viral peptide (from an influenza type A hemagglutinin [black]), recognized by T cell receptor α and β chains (green and dark blue respectively) ¹. This can be compared with Fig. 3.4 of *Searching for Molecular Solutions*, where recognition of peptide bound by MHC Class I is shown (the same colors for T cell receptor α and β chains as for the above Fig. 3.Na are also used in the Fig. 3.4 [Color Version](#) within this ftp site). Source: [Protein Data Bank](#) ² 1J8H. Images generated with Protein Workshop ³.

As noted in *Searching for Molecular Solutions*, the peptide-binding groove for Class II MHC is formed from association between separate α and β chains (human HLA-DR α and DR-4 β respectively, in Fig. 3.Na), rather than as a single chain in the case of Class I MHC (Fig. 3.4).

Another significant difference between the MHC classes lies in the length of bound peptides. Class II MHC molecules typically bind longer peptides than for Class I. For example, in the above Fig. 3.Na, the bound influenza hemagglutinin peptide (PKYVKQNTLKLAT) is 13 residues long, but the Class I-bound peptide of *Searching for Molecular Solutions* Fig. 3.4 is only nine residues in length (Human T cell Leukemia Virus-1 Tax peptide; LLFGYPVYV).

Section 4: ***DNA-Binding Protein Design***

Cited on p. 84 (Footnote) of *Searching for Molecular Solutions*

This section provides some additional background material concerning the engineering of DNA-binding proteins with any desired sequence-specific binding activity, a field in which we can find the application of directed evolution with a blend of rational input, leading towards fully rational design. But for all approaches with this end in mind, structural information is of prime importance.

Although many applications for 'designer' DNA-binding proteins can immediately be put forward, these can be broken down into two overarching categories: design for sequence-specific cleavage or strand nicking, and design for directing a specific functional activity to a predetermined DNA sequence. Examples of the latter include the control of gene expression through sequence-specific targeting of protein domains with activator or repressor functional properties. Sequence-specific cleavage at any desired site (*in vitro* or *in vivo*) has obvious ramifications for general genetic engineering applications, but is also of great interest for directing recombination towards the generation of desired products.

Before proceeding further, we should note that a major class of DNA-binding proteins intensively studied for manipulation of sequence-specific binding are those with zinc finger domains, which will be considered in more detail in a separate *Searching for Molecular Solutions* note (SMS–CitedNotes-Ch4 / Section 8B; from the same ftp site). Here, we will look at some additional specific cases: the design of restriction enzymes and homing endonucleases.

Restriction enzymes by design

In Chapter 3 of *Searching for Molecular Solutions* (p. 64), restriction enzymes were noted as a form of 'bacterial immunity', by virtue of their function of inactivating invading nucleic acids of viral parasites (depicted in Fig. 3.Nb).

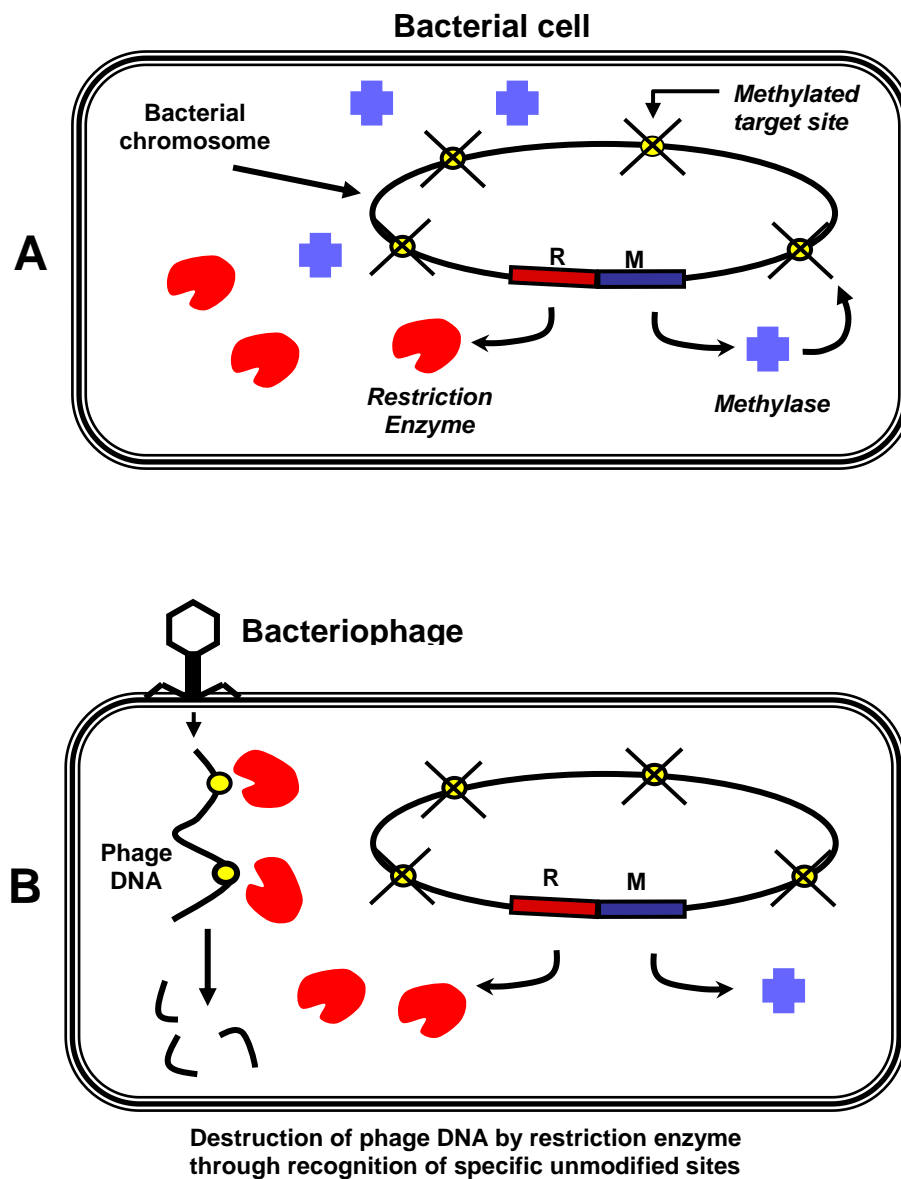


Fig. 3. Nb

Fig. 3.Nb. Depiction of bacterial restriction-modification system for protection against foreign DNA, so named historically for its ability to *restrict* the range of foreign DNAs compatible with specific bacterial strains. **A**, Bacterial cell with a restriction-modification system expressing the restriction enzyme and its accompanying methylase from their corresponding genes (R and M respectively). The methylase enzyme modifies specific target sites (small circles) during replication of the bacterial chromosome, preventing self-damage by the restriction enzyme which recognizes corresponding unmethylated sites. **B**, An invading bacteriophage (phage, or bacterial virus) injects its DNA into the host. The phage DNA contains unmodified sites which are recognized and cleaved by the restriction enzyme, resulting in prevention of phage replication. For some types of restriction enzymes, additional mechanisms beyond methylation exist for protection of host DNA; in effect additional means for discriminating non-self^{4,5}. See Chapter 3 of *Searching for Molecular Solutions* for an extended discussion of self and non-self issues.

The simple schematic depiction of a restriction-modification system in Fig. 3.Nb does not convey the considerable diversity and complexities which have been revealed over decades of investigation. Restriction enzymes are grouped into three major categories (Types I, II, III[♥]) based on their architectures, substrate requirements, and recognition properties; these classes can in turn be split into many subclasses⁶. For molecular biological applications, enzymes of the Type II class which cleave at (or close to) specific sequences have proven the most useful. (Type I enzymes, for example, recognize specific sequences but cleave at variable sites away from the actual recognition target).

By their nature, restriction enzymes have historically been ‘found objects’ in the sense that they are highly useful products of specific prokaryotic organisms in the biosphere, as revealed through human inspection. Ever since their first description and the realization that the diversity of microbial populations was a rich potential source of new enzymes, intensive and systematic screening programs have been instituted towards finding novel DNA sequence recognition

[♥]A fourth class (Type IV; which cleaves only modified DNA targets) is also recognized⁶.

and cleavage specificities. Sifting through nature's restriction enzymatic bounty in the prokaryotic world has revealed hundreds of different recognition specificities, usually in the range of 4-8 bases ⁷. Even so, this hardly covers every possible short sequence, and molecular biologists have long dreamed of altering the natural recognition specificities of restriction enzymes at will, such that virtually any sequence can be targeted.

Restriction enzyme structures have been available for decades ⁸⁻¹⁰ but early hopes that this information would enable directed changes to recognition specificities were not matched by real results. In the Type II enzymes initially used, changes to amino acid residues [♥] contacting specific bases in DNA recognition sequences were usually accompanied by strong activity losses ¹¹. The frequency of this observation suggested that changes to contact residues would inevitably be accompanied by catalytic impairment without greatly affecting specificity ¹². Moreover, even structurally related Type II restriction enzymes binding similar but non-identical DNA sequences can use different contact residues for recognition purposes ^{13,14}. A major source of the 'immutability' of the prototypical Type II enzymes seems to be reducible to the coupling between recognition and cleavage sites. In other words, the sites of DNA binding and DNA cleavage in such enzymes coincide, rendering it a difficult proposition to change specificity while leaving catalysis intact.

Yet other biological examples have shown that these activities are in principle separable, without necessarily overlapping in terms of protein sequence. For example, the enzyme FokI (Type IIA class) has a specific DNA recognition domain in conjunction with another discrete domain which mediates non-sequence specific strand cleavage at a defined distance from the recognition site. This has been extensively used to create chimeric molecules which combine the cleavage domain with a different DNA recognition motif from a variety of

[♥]This has been the consensus finding whether amino acid changes resulted from directed evolutionary approaches or 'rational' targeting ¹¹.

natural sources^{♥ 16}. Type IIB enzymes have bipartite recognition sequences, and cleave on both sides of their bipartite targets by means of distinct and separable protein domains. This feature has proven useful by allowing the combinatorial shuffling of recognition domains for half-sites of bipartite targets from different Type IIB family members, with the resulting generation of novel hybrid specificities¹⁷.

An enzyme shown to be a fruitful target for the engineering of altered specificity is Mmel, originally derived from the bacterium *Methylophilus methylotrophus*¹⁸. This endonuclease has some unusual properties, leading to the proposal that it (and related enzymes) should be relegated into a separate Type II subclass (Type IIL¹⁹). Not only does it have restriction and methylation-modification activities present within the same polypeptide (already a known feature of Type IIG enzymes^{* 6}, but methylation only occurs on one strand²⁰. As with FokI, Mmel and its kin cleave DNA at specific points beyond the recognition site, although Mmel cleavage occurs at a longer distance (20 base pairs²⁰).

In a demonstration of the power of genomic approaches when many fully sequenced microbial genomes are available, the sequence of the Mmel gene allowed rapid definition of a panel of Mmel homologs (fellow members of the proposed Type IIL restriction enzyme subclass). Analysis of this set demonstrated clear correlations between protein sequence elements and DNA recognition specificities, such that specific amino acid residues could be assigned to specific bases with the Type IIL recognition sequences. With this

♥One such example are the zinc finger nucleases, which is briefly noted in Section 8B of the Chapter 4 cited notes for *Searching for Molecular Solutions* from the same ftp site. As well as sequence-specific recognition motifs, protein domains which recognize and bind alternative duplex DNA structure (Z-DNA) can also be fused with FokI¹⁵. In this case, cleavage occurs at boundaries between Z- and conventional B-DNA structural transitions.

*This is unusual within restriction enzymes as a whole. The situation depicted in Fig. 3.Nb with separate restriction and modification activities is the observed arrangement in the majority of cases.

information in hand, it was shown that combinatorial reassortment of amino acid residues at specific positions changed the recognized DNA sites accordingly ♥¹¹.

It has been noted that Type IIL enzymes have properties suggesting that they are 'naturally designed' for specificity changes through recombination²⁰. In 'conventional' Type II enzymes where restriction (cleavage) and modification (methylation) activities are mediated by separate proteins (as in Fig. 3.Nb), a change in cleavage specificity alone will be lethal, since the methylase will continue to act only upon the 'old' target sequence, leaving accessible all novel target sites within the host genome itself. A specificity switch in conventional Type II enzymes therefore requires parallel mutation in both restriction and modification genes, a very low-frequency event. Yet for MmeI and Type IIL enzymes, restriction and modification activities are present within a single protein and uncoupled from the recognition domain. In other words, since both cleavage and methylation activities for MmeI are directed to specific target sites through a shared protein region mediating DNA sequence recognition, changes in such sequence recognition automatically produce the required alterations in restriction and modification. This versatility is reflected in the utility of MmeI for engineering of specificity alterations, but the ability of this enzyme to accommodate a wide range of amino acid substitutions without compromising catalysis suggests that it is also highly flexible, at least in its recognition domain. Overall, this combination of properties has rendered MmeI the most successfully engineered restriction enzyme to date.

♥ For example, a subset of Type IIL enzymes (including MmeI itself) have C at position 6 of their DNA recognition sequences, and this was associated with a strong preference for amino acids E and R at specific and equivalent positions in their protein sequences. Likewise, another Type IIL subset with a DNA recognition site bearing G at position 6 have preferences for amino acids K and D at the corresponding protein sequence positions. Swapping the E/R and K/D residues also swaps the sixth-base C-G recognition specificity for DNA target sequences¹¹.

Before moving on from this area, the recent elegant work with Mmel is also interesting when viewed through the empirical-rational discovery theme of *Searching for Molecular Solutions*. This work has been termed 'rational engineering'¹¹, which can be distinguished from 'rational design' in its strongest sense. In particular, the Mmel engineering referred to here was not based on structural information (which was not available when the work was performed), but rather on sophisticated bioinformatics and sequence alignments. Even if an absolute correlation exists between specific positions within recognized DNA sequences and specific amino acid residues at the protein level[▼], without further information it does not necessarily follow that novel recognized DNA targets will be accordingly generated in mix-and-match experiments, since structural constraints often impede such a cross-over of functions.

As noted above, the evident remarkable flexibility of Mmel has promoted favorable outcomes to the protein engineering experiments, but this was a fortuitous finding not predicted in advance through structural information. Initial experimental findings suggesting the tolerance of Mmel recognition for sequence shuffling then increase the level of confidence for subsequent predictions of new specificities based on sequence-function information. Yet although analyses of Type III enzymes have enabled successful recognition specificity predictions based on amino acid residue combinatorics, without structural information the understanding of the process is incomplete, and therefore cannot be considered 'rational' in the strongest sense. In the empirical-rational spectrum discussed within *Searching for Molecular Solutions* (principally in Chapter 9), this kind of protein engineering is therefore still within a gray area often termed 'semi-rational' design. Simply from a strategic engineering point of view, though, the pathway taken for this work is highly logical, and thus the term 'rational engineering' itself is quite reasonable.

[▼]Most of the reported correlations followed strong patterns, but were not absolutely adhered in all observed cases¹¹.

Homing in on Homing Endonucleases and Other Things

As noted, the specific DNA sequence recognition targets for restriction enzymes are typically 4-8 bases long ♥, but this is by no means the biological limit of such recognition. Numerous enzymes termed ‘homing endonucleases’ recognize much larger sequences (from 14 to over 40 bp, although with tolerance for individual base changes within the recognized sites ²¹⁻²³) and are as a result often termed ‘megannucleases’. Before considering work on changing their target specificities, some background material is in order.

The *raison d’être* of homing endonucleases is to promote their own transmission to new specific target sites, along with accompanying sequences. As such, they act as quintessential ‘selfish’ DNA elements promoting their own survival. The mechanism by which such endonucleases accomplish this is simple enough: they recognize target sites in homologous genes previously free of their own coding sequences and cleave them. In other words, their primary targets are ‘uninfected’ copies of the local host genomic regions which carry their coding sequences in ‘infected’ hosts ²². Following double-stranded cleavage, exonucleolytic action on the DNA ends exposes 3’ single-stranded regions which promote the form of directionalized homologous recombination commonly known as gene conversion. The end result is indeed conversion of the ‘uninfected’ gene into a copy bearing the endonuclease, and the transmission has succeeded (Fig. 3. Nc). But if one thinks about this process, some problems become apparent (or at least, problems from the blind and imaginary ‘point of view’ of the

♥Note the reference to *specific* bases, in that specified bases within a recognition site can be interspersed with non-specific ones. For example, the enzyme SfiI recognizes the sequence GGCCNNNN↓NGGCC, and FseI recognizes GGCCGG↓CC (arrows indicate cleavage sites; from the REbase database run by [New England Biolabs](#)). Since both recognize 8 specific bases, they have the same frequency of occurrence in a long random DNA sequence (4^8 , or approximately once every 65 kilobases). But SfiI requires five unspecified bases between the two GGCC motifs to enable recognition to occur.

endonuclease as a mobile element). If an endonuclease inserts itself alone into all copies of essential gene (two for a diploid organism) then the gene is inactivated and the host is killed, defeating the purpose of maintaining the transmission of the selfish coding element. Yet if non-essential genes or dispensable intergenic regions are targeted, the parasitic element can be rapidly lost by mutation. Although 'naked' homing endonuclease gene coding sequences acting as isolated mobile elements have been identified²⁴, this is only applicable in special circumstances.

The selfish solution is to come up with a mechanism whereby the endonuclease can insert itself into essential genes *without* harming the host. There are two ways, at quite different levels, by which this can be done. Firstly, since some intervening sequences in genes can be removed by self-splicing at the RNA level, insertion of an endonuclease coding sequence entirely within such an intron will likewise remove it from mature mRNAs, allowing normal expression to continue (Fig. 3. Nc). The alternative is to use an analogous process at the protein level. Peptide sequences which mediate splicing of protein intervening segments or *inteins*[♥] exist, and embedding of an expressed endonuclease within an intein will also remove it from the final active protein sequence (Fig. 3. Nc).

[♥]A little more background about inteins is provided in the Extra Material for Chapter 3 (SMS-Extras-Ch3/Section A2) from the same ftp site.

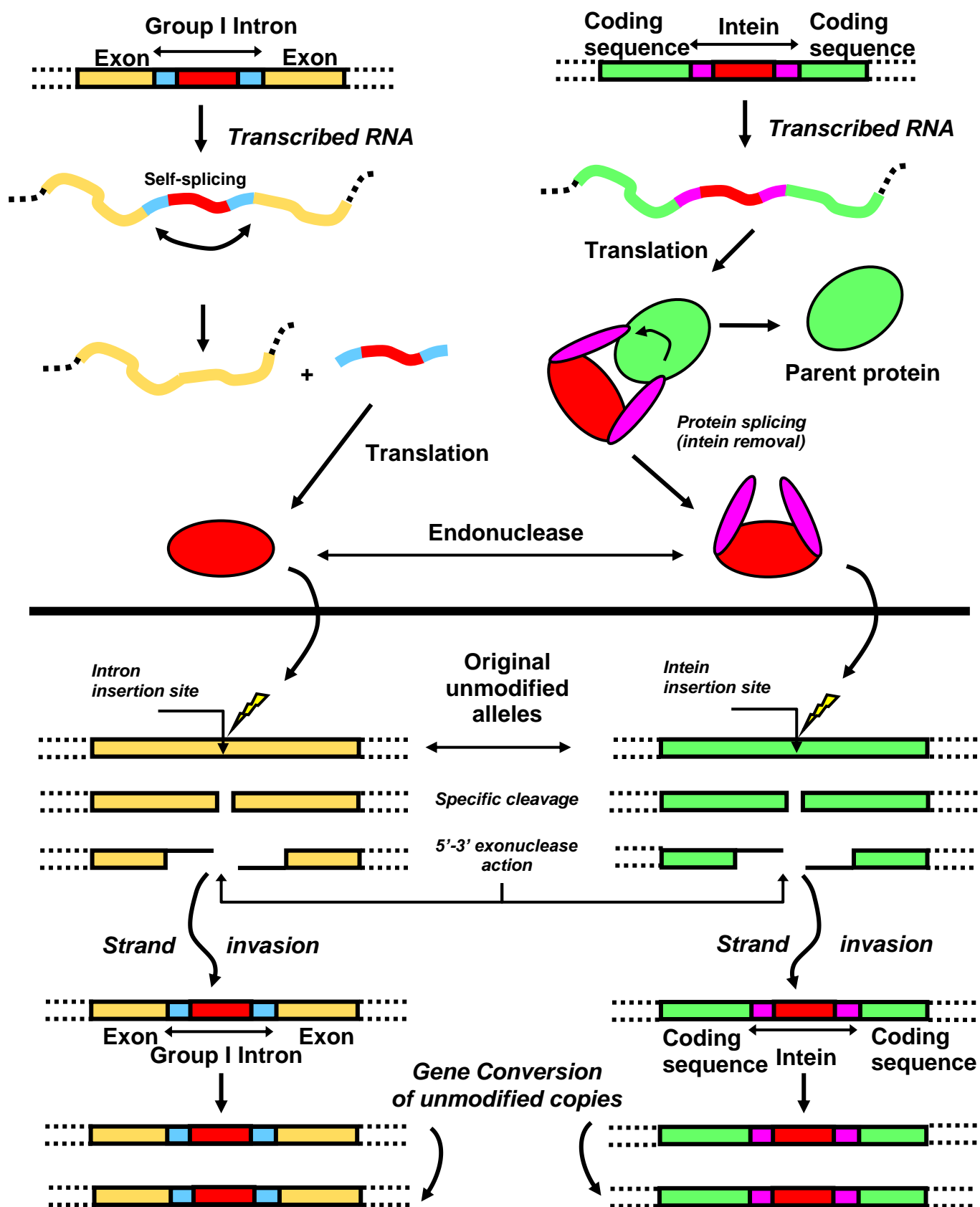


Fig. 3. Nc

Fig. 3. Nc. Mechanisms of homing endonucleases as mobile elements, or their ‘homing’ activities within Group I introns and inteins ²². In the top panel, the generation of active endonucleases from their coding sequences embedded within Group I self-splicing introns or protein inteins is depicted. For the former, the intronic RNA containing the endonuclease coding sequence is self-spliced from its parental mRNA, and subsequently translated to yield the endonuclease protein. In the intein circumstance, the parent protein is expressed and then the endonuclease / intein sequence is removed through a protein splicing reaction. (Note that the intein peptide sequences flanking the endonuclease are included in the spliced-out polypeptide sequence; the intein residue is often co-opted as part of the endonuclease binding site itself ²⁵). The bottom panel then depicts the process of ‘homing’ itself, which is initiated by a double-stranded cleavage event mediated by the exonucleases at corresponding homologous unmodified gene alleles. (Following transposition of the intron or intein sequence, the recognition site is disrupted, such that the modified allele cannot be re-targeted by the same endonuclease). Accessory enzymes then allow the type of homologous recombination termed gene conversion to take place. Although multiple alternatives exist, one pathway uses 5' → 3' exonucleases for generation of long single-stranded 3' sequences which effect strand invasion of the modified allele. Following strand extension and resolution, two copies of the modified allele result, and copying of the intron or intein into the virgin allele. Other complex homing mechanisms with Group II introns also exist ²².

The ‘needs’ of homing endonucleases as mobile elements also explain why they characteristically recognize such long sequences, especially in comparison with restriction enzymes. The homing enzymes target sites within specific unmodified alleles, and bystander cutting at spurious unrelated sites would not only be fruitless for transmission by homologous recombination, but quite deleterious for their hosts. Therefore, evolutionary pressures favor highly selective DNA target choice, towards only one recognition site per genome. For restriction enzymes, on the other hand, as long as the host genome is protected by methylation, a

short recognition sequence found at high frequencies within invading DNA molecules is an effective strategy.

Continuing these kinds of comparisons, it would seem at first glance that restriction enzymes and homing endonucleases have evolved from fundamentally different directions. The latter may appear as 'bad' selfish parasites, while the former act as 'good' defenders of their host genomes. While such an anthropomorphic stance is inherently misleading, in any case this simple dichotomy itself breaks down upon deeper examination. Protein folds used by restriction enzymes can be found in homing endonucleases ²⁶, and there are good reasons for viewing restriction / modification systems as 'selfish' in their own right ²⁷, even though they confer an obvious survival benefit to their hosts ♥. Indeed, it has been possible to create artificial mobile elements using a restriction enzyme as the mediating nuclease ²⁸. Conversely, there are recorded circumstances where initially parasitic homing endonucleases have become 'domesticated' through their evolution of useful roles for host fitness ²².

The interest in engineering of homing endonucleases is driven by a number of factors, but major considerations are their long recognition sites and extremely low cutting frequencies for random DNA, which render them as attractive prospects for many genomic engineering tasks. In particular, double-stranded breaks created by these enzymes can be used to stimulate homologous recombination for gene targeting purposes *. If an enzyme can be engineered to cleave a unique sequence at a genomic site of interest, a powerful adjunct to current gene targeting technology emerges by inducing gene conversion events

♥For example, it is difficult (although not impossible) for a host bacterium to lose a restriction-modification once it has acquired it. If the restriction-modification genes are lost, inactivated, or blocked, newly replicated host DNA becomes sensitive to the original restriction enzyme, and any residual persisting enzyme can then destroy the unmodified new host DNA ²⁷.

*This refers to the modification of genomic sites in a pre-determined manner, very often in embryonal stem cells for the purposes of generating mice with specific genetic changes.

²⁹ in a similar manner as portrayed in Fig. 3.Nc. Another strong incentive for the development of homing endonucleases with 'designer' target site recognition is for gene therapy ³⁰, which is analogous to related applications envisaged for zinc-finger nucleases (see SMS–CitedNotes-Ch4 / Section 8B, from the same ftp site). Induced gene conversion events for such applications are designed to replace an existing defective genetic allele with a corrected version.

Although at least four distinct structural classes of homing endonucleases have been defined ²², most work towards the redesign of specificity has used the 'LADLIDADG' class of enzymes (so named from a recurring sequence motif in this group). As with classical Type II restriction enzymes, the cleavage and recognition domains in LADLIDADG homing endonucleases are coupled, but there has been much more success in redesigning the latter for specificity alterations than for Type II restriction enzymes themselves (notwithstanding recent advances in this field for Type III enzymes as noted above). In general, this is linked with the longer recognition sites for homing endonucleases and their tolerance for single-base degeneracies, but a very practical advantage has been the development of effective *in vivo* screening methods for new specificities. Although a number of these have been utilized, they all rely on the primary role of the double-stranded breaks generated by homing endonucleases in promoting homologous recombination ^{22,31,32}.

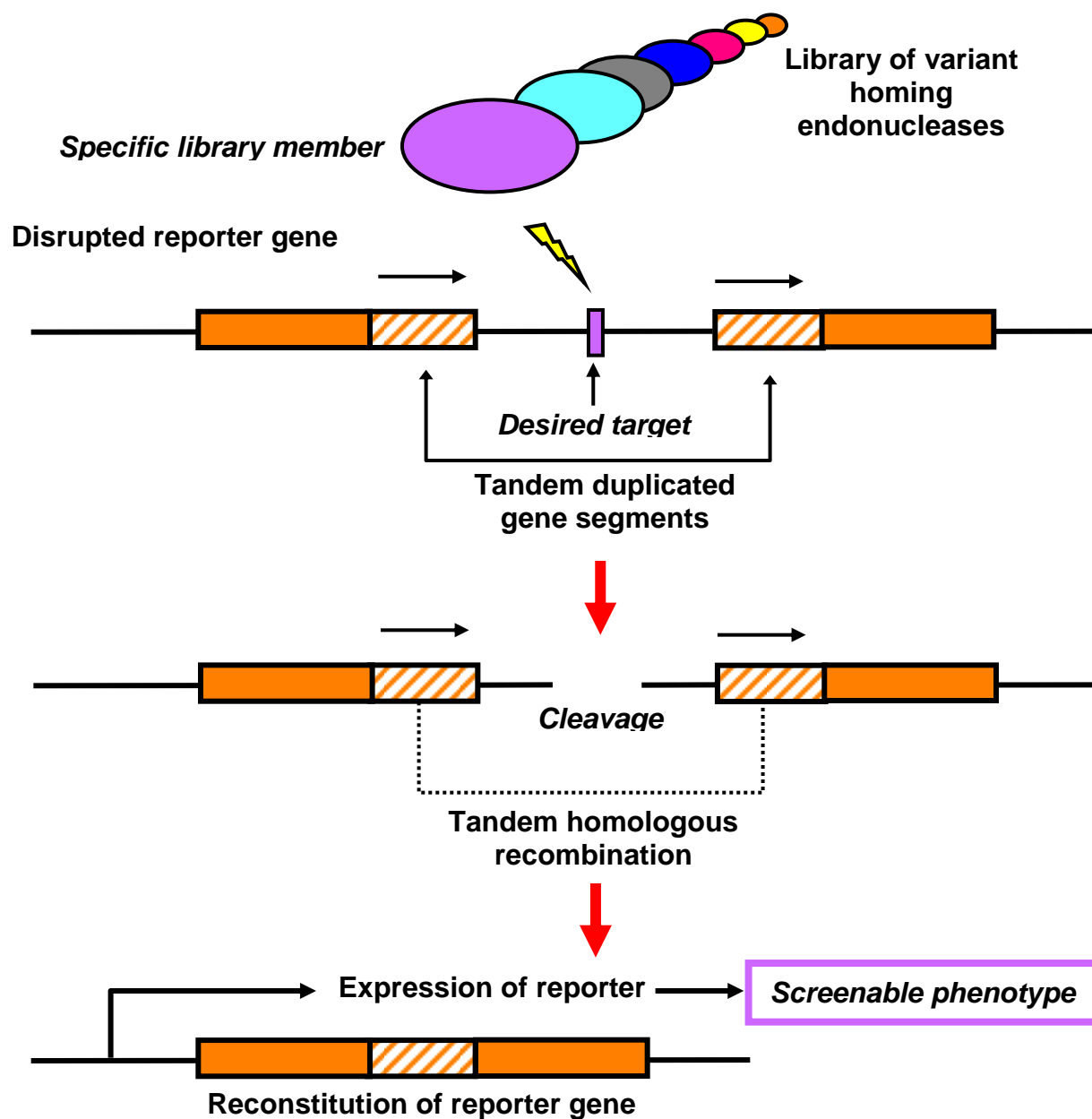


Fig. 3.Nd

Depiction of a system enabling phenotypic screening for a new recognition and cleavage specificity from a library of variant homing endonucleases (typically expressed in yeast cells). A reporter is engineered such that segments of it are tandemly duplicated and flank an inserted sequence which disrupts its correct expression, and which also contains the desired recognition site (altered with respect to the naturally-recognized site). A member of the library capable of cleaving this site will generate a double-stranded break and promote tandem homologous recombination which restores the

reporter gene and allows a phenotypic read-out. (For example, the *lacZ* gene allows screening for the generation of color by means of chromogenic substrates for the expressed β -galactosidase enzyme).

The size of homing endonuclease recognition sites in itself is an advantage for this screening application, since their very low frequencies of occurrence in random DNA renders it unlikely that genomes of experimental organisms will have spurious cleavage sites [▼]. (This is directly testable with knowledge of the complete genomes of yeast [*Saccharomyces cerevisiae*] and *E. coli*). An example of a system enabling effective screening for a new variant recognition and cleavage specificity is depicted in Fig. 3.Nd. Selection for novel binding specificities can also be approached by phage display ³³, in an analogous manner as for zinc finger proteins (see SMS–CitedNotes-Ch4 / Section 8B, from the same ftp site).

The specific DNA cleavage event engendered by homing endonuclease can be also adapted into a positive selection system, if the cleavage event is coupled with cell survival. This can be achieved in *E. coli* systems by placing a desired target site adjacent to a toxic gene on an independently-replicating plasmid from that encoding the library of homing nucleases themselves. Only cells which can cleave the target site (that is, cells which express a variant nuclease with the appropriate specificity) can therefore theoretically survive ³⁴⁻³⁶.

Much work aimed at tinkering with homing endonuclease cleavage specificities can be classed as semirational. Structural information which provides insights into specific contacts between recognition sites and enzymes enables highly rational decisions to be made regarding choices of enzyme residues to alter for

[▼]One could not, for example, use an analogous strategy for screening for novel 4- or 6-base restriction enzyme specificities, since many sites of this relatively short length (and much higher frequencies of random occurrence) would be found scattered throughout the host genome.

changing recognition specificities. Randomization of such residues of interest enables the generation of directed enzyme libraries of limited size, and their screening or selection through processes outlined above. Many variants of the parental recognition sequences can be obtained in such a fashion^{31,36}. In addition, studies of LADLIDADG enzymes collectively have allowed domain swapping between different family members, with concomitant diversification of sequence recognition³⁷. Domain assortment can be also carried out in an analogous manner as for the chimeric restriction enzymes referred to above, and in fact also by means of the useful separation between the recognition and cleavage domains of the Type IIA enzyme FokI. The non-specific FokI cleavage domain can thus be linked with the recognition domain of a LADLIDADG homing enzyme (I-SceI) where the latter is inactivated for its own endonuclease activity³⁸.

Initial homing endonuclease variants found by the above screening or selection processes tend to recognize local divergences from the parental target sites rather than radically different sequences, but this primary work can act to bootstrap subsequent modifications. Thus, a large set of variant enzymes identified by a primary screening process³¹ enabled statistical definition of an amino acid residue / DNA contact code, and the logical generation of further recognition diversity through combinatorial shuffling³².

Accumulation of structural information for increasing numbers of LAGLIDADG enzymes improves prospects for rational design²³. Computational remodeling using Rosetta software (described in more detail in Chapter 9 of *Searching for Molecular Solutions*) has been applied to a member of this class of homing endonucleases, with a successful specificity change for 2/24 bases in the wild-type recognition sequence, and a change in two residues in the endonuclease protein³⁹.

For real-world applications, a typical need is to find a homing enzyme site within a target gene of interest in its normal genomic context. Since the extremely low frequency of long recognition sites in random DNA renders the chances of finding such a natural site vanishingly small, the best option is to search for a site as close as possible to a given homing endonuclease, and then mutationally adapt the enzyme for recognition of the new desired site. Artificially-engineered double-stranded breaks in genes of interest can then allow directed homologous recombinational events, a springboard for gene therapy. Successful targeted recombinational gene corrections have been reported, initiated through specific DNA cleavage by modified homing endonucleases⁴⁰⁻⁴², and these promising early results auger well for future gene therapeutic applications. Such engineered nucleases are generally regarded as having significantly lower chances of inducing non-specific strand breaks than the competing technology of zinc-finger nucleases^{22,39}.

Homing endonucleases stand as example of a natural parasitic element which can be highly valuable as a molecular tool (or as a source of material for subsequent tool development) for humans. This can even be regarded as a higher-level case of the 'domestication' of such inherently parasitic elements by their hosts⁴³, as note above. In an even more direct example of 'getting something back from a parasitic entity', some bacterial species which have historically been major afflictions of humanity have donated useful restriction enzymes. One example is Styl from *Salmonella typhi* (the causative agent of typhoid fever). We might reflect that the royal doctors who unsuccessfully attempted to save the life of the notable typhoid victim Prince Albert (consort of Queen Victoria) would no doubt have considered as ridiculous the notion that the source of such an affliction could have any possible useful application.

In fact, genomic approaches have allowed the identification of at least putative restriction and / or methylation enzymes, and sometimes homing endonucleases, from wide variety of pathogenic bacteria, including *Streptococcus pyogenes*,

Brucella abortus, *Francisella tularensis*, *Mycobacterium tuberculosis*, and *Yersinia pestis* ([REbase](#)). These are the causative agents of scarlet fever, brucellosis, tularemia, tuberculosis, and bubonic plague, respectively. The last two cases in particular give one pause to consider that it would indeed be a nice thought if such incredibly destructive organisms could at least offer something useful back to humanity.

Yet unfortunately, perhaps, this principle does not extend too far. As the number of known natural recognition specificities increases, a point of diminishing returns is reached as it becomes harder and harder to find new ones, especially for short sequences ♣. In other words, casting a wider screening net for novel enzymes tends to yield reduced amounts of true novelty per time spent. Yet in a sense, it may soon be said that natural sources (whether eukaryotic parasitic self-replicating elements or prokaryotic self-protection systems) have already done their job of providing not only the physical enzymatic tools for molecular biology, but the information needed to enable the generalization of design for all future such tools. In historical terms, there will be a very narrow window of time between the era when enzymes cleaving DNAs at specific sites were obtained from the biological environment, and when they were routinely engineered towards pre-chosen specificities.

In concluding this section, it will be apparent that we have concentrated on a special subset of DNA-binding proteins: those which also cleave DNA. As we have seen, cleavage and DNA recognition and binding are quite distinct, and of course a vast number of natural proteins (especially transcription factors) have roles which call for specific DNA binding only. Yet the dichotomy between cleavage and binding can be artificially bridged from both directions.

♣ It should be noted though, that bacterial sources can also yield homing nucleases. Thus in at least one of the above examples (*M. tuberculosis*), a novel homing endonuclease with a long recognition sequence has been well-defined ⁴⁴.

Transcription factor binding sites can be harnessed for nuclease activity, as with zinc finger proteins (as noted above; SMS–CitedNotes-Ch4 / Section 8B; from the same ftp site), and nucleases can be modified to perform sequence-specific DNA-binding only ³⁸. The diversity of DNA-binding proteins in general ⁴⁵ suggests that a diversity of engineering approaches and diverse screening or selection strategies will be required. Certainly non-cleaving DNA-binding proteins obviously cannot be screened or selected for through some of the ingenious methods devised for homing endonucleases above. On the other hand, with structural information in hand, rational design methods developed for homing enzymes should find much more broad applicability within generalized DNA-binding proteins.

A final take-away thought is the importance of protein domain modularity in so many of the above studies. This theme is extended further in the brief description of the two-hybrid system and its relatives, also in Cited Notes for *Searching for Molecular Solutions* Chapter 4, Section 9, from the same ftp site.

Section 5: **Neural Diversity**

Cited on p. 92 of *Searching for Molecular Solutions*

Once the somatic genomic DNA rearrangements of the adaptive immune system were understood, it was a worthwhile objective to search for analogous events within central nervous system neurons, given certain parallels between the immune and neural systems[▼]. The overall position reached from a number of such studies has been that neural somatic DNA recombination does not occur in the same manner as in the immune system⁴⁶. This conclusion in itself does not rule out the possibility that some other independent form of somatic genomic rearrangements could occur with the nervous system. Some intriguing studies of possible relevance in this regard have been performed with proteins involved with the non-homologous end-joining of double-stranded DNA breaks, DNA ligase IV and its accessory protein XRCC4. Mice with either of these genes artificially 'knocked out' suffer both impaired V(D)J immune system recombination and extensive neuronal cell death leading to embryonic lethality⁴⁷. A more recent report of neural somatic DNA recombination has been shown to be not specific for brain tissue, and its general significance remains unclear^{48,49}. Another intriguing suggestion is that transposable elements (especially retrotransposons) may be mediators of neural somatic diversity^{50,51}. Once again, the generalizable significance of this has yet to be pinned down.

So somatic changes at the DNA level, irrespective of parallels with the immune system, have not yet been linked with neural diversity. And yet, from logical consideration of both higher-level and cellular neural processes, somatic diversification mechanisms at some level must exist. Complex neural wiring is a process (obviously specific to the nervous system) which requires great feats of

▼ One such parallel noted in *Searching for Molecular Solutions* is the existence of cellular synapses in both systems. Further details on immune / neural analogies are provided in the Extra material for Chapter 3 (SMS-Extras-Ch3/Section A4; also from the same ftp site).

recognition specificity, and implicit high molecular diversity of some form or another. (The Nobel Prize winner Roger Sperry proposed in 1963 that such wiring was mediated by 'chemoaffinity tags'⁵²). The immune system itself teaches us that genomic DNA rearrangements are by no means the only potential pathway towards achieving the required somatic diversity. As another example, in the fruitfly nervous system there is a precedent for high levels of such diversity with alternate splicing mechanisms; specifically in the case of DSCAM, which has the potential for 38,016 variants ♥^{54,55}. Although this does not appear to be the pathway used by vertebrates, alternative diverse recognition-related molecules such as the cadherin family have been proposed⁵⁶⁻⁵⁸. RNA editing, another means for diversification of genomic information (noted in this Chapter 3 of *Searching for Molecular Solutions* in the context of innate immunity), is known to occur extensively in neural systems. Fruitflies and nematode worms with inactivated RNA editing enzymes are viable but show behavioral defects, and mice with equivalent engineered mutations cannot grow to maturity⁵⁹. A number of central nervous system targets for RNA editing have been defined⁵⁹.

As an add-on to the central theme of the immune system, this account of neural somatic diversity is necessarily brief and not intended to be a full rendition of the state of the art. Nevertheless, it also carries the implicit message that there is a very large amount of information in this field which is still to be gained. (This is not to suggest by any means that research immunologists are in imminent danger of unemployment, but it is safe to say that immune systems are better understood than neural systems from a global point of view. This is probably in accordance with the opinion that the latter systems are of the most supreme complexity). By the same token, doubtless many messages from neural systems concerning molecular and supramolecular diversification mechanisms remain to be deciphered. Although some interesting parallels do exist between the immune

♥ An interesting aside, and another parallel between neural and immune systems in general, is that high diversity in *Drosophila* DSCAM is also exploited by the fruitfly's innate immune system

and nervous systems (in some aspects they could be regarded as 'sister systems' ♥), evidence to do date suggests they differ in the precise mechanisms by which they achieve somatic diversity.

♥ Again, see SMS-Extras-Ch3/Section A4; also from the same ftp site.

References:

1. Hennecke, J. & Wiley, D. C. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J Exp Med* **195**, 571-81 (2002).
2. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980 (2003).
3. Moreland, J. L., Gramada, A., Buzko, O. V., Zhang, Q. & Bourne, P. E. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**, 21 (2005).
4. Makovets, S., Doronina, V. A. & Murray, N. E. Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc Natl Acad Sci U S A* **96**, 9757-62 (1999).
5. Loenen, W. A. Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res* **31**, 7059-69 (2003).
6. Roberts, R. J. et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**, 1805-12 (2003).
7. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* **62**, 685-707 (2005).
8. Frederick, C. A. et al. Kinked DNA in crystalline complex with EcoRI endonuclease. *Nature* **309**, 327-31 (1984).
9. Jen-Jacobson, L., Lesser, D. & Kurpiewski, M. The enfolding arms of EcoRI endonuclease: role in DNA binding and cleavage. *Cell* **45**, 619-29 (1986).
10. McClarin, J. A. et al. Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å resolution. *Science* **234**, 1526-41 (1986).
11. Morgan, R. D. & Luyten, Y. A. Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Res* (2009).
12. Alves, J. et al. Changing the hydrogen-bonding potential in the DNA binding site of EcoRI by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site-GAATTC. *Biochemistry* **28**, 2678-84 (1989).
13. Lukacs, C. M., Kucera, R., Schildkraut, I. & Aggarwal, A. K. Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 Å resolution. *Nat Struct Biol* **7**, 134-40 (2000).
14. Lukacs, C. M. & Aggarwal, A. K. BglII and MunI: what a difference a base makes. *Curr Opin Struct Biol* **11**, 14-8 (2001).

15. Kim, Y. G., Kim, P. S., Herbert, A. & Rich, A. Construction of a Z-DNA-specific restriction endonuclease. *Proc Natl Acad Sci U S A* **94**, 12875-9 (1997).
16. Chandrasegaran, S. & Smith, J. Chimeric restriction enzymes: what is next? *Biol Chem* **380**, 841-8 (1999).
17. Jurenaite-Urbanaviciene, S. et al. Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc Natl Acad Sci U S A* **104**, 10358-63 (2007).
18. Boyd, A. C., Charles, I. G., Keyte, J. W. & Brammar, W. J. Isolation and computer-aided characterization of Mmel, a type II restriction endonuclease from *Methylophilus methylotrophus*. *Nucleic Acids Res* **14**, 5255-74 (1986).
19. Morgan, R. D., Dwinell, E. A., Bhatia, T. K., Lang, E. M. & Luyten, Y. A. The Mmel family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res* (2009).
20. Morgan, R. D., Bhatia, T. K., Lovasco, L. & Davis, T. B. Mmel: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res* **36**, 6558-70 (2008).
21. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J., Jr. & Stoddard, B. L. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-Crel and I-MsoI. *J Mol Biol* **329**, 253-69 (2003).
22. Stoddard, B. L. Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49-95 (2005).
23. Marcaida, M. J. et al. Crystal structure of I-Dmol in complex with its target DNA provides new insights into meganuclease engineering. *Proc Natl Acad Sci U S A* **105**, 16888-93 (2008).
24. Belle, A., Landthaler, M. & Shub, D. A. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev* **16**, 351-62 (2002).
25. Moure, C. M., Gimble, F. S. & Quijcho, F. A. Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nat Struct Biol* **9**, 764-70 (2002).
26. Zhao, L., Bonocora, R. P., Shub, D. A. & Stoddard, B. L. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *Embo J* **26**, 2432-42 (2007).
27. Kobayashi, I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* **29**, 3742-56 (2001).
28. Eddy, S. R. & Gold, L. Artificial mobile DNA element constructed from the EcoRI endonuclease gene. *Proc Natl Acad Sci U S A* **89**, 1544-7 (1992).

29. Cabaniols, J. P. & Paques, F. Robust cell line development using meganucleases. *Methods Mol Biol* **435**, 31-45 (2008).
30. Paques, F. & Duchateau, P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**, 49-66 (2007).
31. Arnould, S. et al. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* **355**, 443-58 (2006).
32. Smith, J. et al. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* **34**, e149 (2006).
33. Chames, P. et al. In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res* **33**, e178 (2005).
34. Gruen, M., Chang, K., Serbanescu, I. & Liu, D. R. An in vivo selection system for homing endonuclease activity. *Nucleic Acids Res* **30**, e29 (2002).
35. Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* **33**, e154 (2005).
36. Chen, Z., Wen, F., Sun, N. & Zhao, H. Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng Des Sel* **22**, 249-56 (2009).
37. Epinat, J. C. et al. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res* **31**, 2952-62 (2003).
38. Lippow, S. M. et al. Creation of a type IIS restriction endonuclease with a long recognition sequence. *Nucleic Acids Res* **37**, 3061-73 (2009).
39. Ashworth, J. et al. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656-9 (2006).
40. Gouble, A. et al. Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. *J Gene Med* **8**, 616-22 (2006).
41. Arnould, S. et al. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol* **371**, 49-65 (2007).
42. Grizot, S. et al. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res* (2009).
43. Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* **14**, 609-15 (2004).
44. Guhan, N. & Muniyappa, K. Mycobacterium tuberculosis RecA intein, a LAGLIDADG homing endonuclease, displays Mn(2+) and DNA-dependent ATPase activity. *Nucleic Acids Res* **31**, 4184-91 (2003).
45. Garvie, C. W. & Wolberger, C. Recognition of specific DNA sequences. *Mol Cell* **8**, 937-46 (2001).

46. Schatz, D. G. & Chun, J. J. V(D)J recombination and the transgenic brain blues. *New Biol* **4**, 188-96 (1992).
47. Gao, Y. et al. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891-902 (1998).
48. Maeda, T. et al. Somatic DNA recombination yielding circular DNA and deletion of a genomic region in embryonic brain. *Biochem Biophys Res Commun* **319**, 1117-23 (2004).
49. Maeda, T. et al. Somatic DNA recombination in a mouse genomic region, BC-1, in brain and non-brain tissue. *Can J Physiol Pharmacol* **84**, 443-9 (2006).
50. Muotri, A. R. et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903-10 (2005).
51. Muotri, A. R., Marchetto, M. C., Coufal, N. G. & Gage, F. H. The necessary junk: new functions for transposable elements. *Hum Mol Genet* **16 Spec No. 2**, R159-67 (2007).
52. Sperry, R. W. Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns and Connections. *Proc Natl Acad Sci U S A* **50**, 703-10 (1963).
53. Watson, F. L. et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* **309**, 1874-8 (2005).
54. Schmucker, D. et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-84 (2000).
55. Schmucker, D. & Flanagan, J. G. Generation of recognition diversity in the nervous system. *Neuron* **44**, 219-22 (2004).
56. Wu, Q. & Maniatis, T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**, 779-90 (1999).
57. Wu, Q. Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics* **169**, 2179-88 (2005).
58. Zipursky, S. L., Wojtowicz, W. M. & Hattori, D. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem Sci* **31**, 581-8 (2006).
59. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **71**, 817-46 (2002).