Combinatorial alanine-scanning
Kim L Morrison and Gregory A Weiss*

Combinatorial libraries of alanine-substituted proteins can be used to rapidly identify residues important for protein function, stability and shape. Each alanine substitution examines the contribution of an individual amino acid sidechain to the functionality of the protein. The recently described method of shotgun scanning uses phage-displayed libraries of alanine-substituted proteins for high-throughput analysis.

Addresses
Department of Chemistry, University of California, Irvine, CA 92697-2025, USA
*e-mail: gweiss@uci.edu

1367-5931/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations
hGH human growth hormone
hGHbp hGH-binding protein

Introduction
Combinatorial protein libraries offer powerful tools to dissect biological phenomena. Protein libraries can be used to explore the intricate relationship between the primary amino acid sequence and protein shape, stability and activity. Mutagenesis of specific residues in proteins has proven invaluable in probing the contributions of individual amino acid sidechains to the properties of proteins. The seminal chemical syntheses of proteins with modified sidechains by Merrifield and co-workers (for a representative example, see [1]), for example, elucidated the enzymatic activity of RNase A. Fersht, Winter and co-workers [2] elegantly demonstrated the power of oligonucleotide-directed mutagenesis to measure hydrogen-bond strengths in proteins.

Combinatorial libraries of mutant proteins can be used to rapidly analyze protein function and identify important sidechain functionalities. For the examination of receptor–ligand interactions, protein libraries can be generated with either random [3,4,5•] or site-specific mutations. Saturation mutagenesis substitutes specific positions with all 20 naturally occurring amino acids [6,7]. However, substitution of a few carefully chosen amino acids can provide a thorough portrait of protein function and simplify analysis of the data. This review focuses upon combinatorial protein libraries substituted with either alanine or the wild-type amino acid in specific positions. Wild type in this context refers to leaving the residue unchanged as the naturally occurring amino acid sidechain. Because this topic has not been reviewed before in this journal, the time period reviewed here is longer than usual.

Non-covalent binding interactions between receptors and ligands are mediated by residues in direct contact, which form a structural epitope. Biophysical techniques including X-ray crystallography and multidimensional NMR reveal such binding contacts. The structure alone, however, does not tell the whole story of how a protein works. Within the contact area, a subset of residues may contribute the majority of binding energy, through hydrogen bonds, salt bridges, dipole–dipole interactions and hydrophobic interactions. Residues with energetically favorable contacts define the functional epitope [8]. A tightly clustered functional epitope resembling a cross section of protein structure (i.e. with hydrophobic center surrounded by hydrophilic

Figure 1

Combinatorial libraries of alanine substitutions. Through site-directed mutagenesis, multiple alanine substitutions are introduced into a wild-type protein. The protein shown here is an α-helical segment of a larger protein, but, in principle, any protein can be targeted for combinatorial alanine scanning. The result is a combinatorial library of proteins with alanine (R = CH₃) or the wild-type sidechain substituted in every position. Because of degeneracy in the genetic code, conventional mutagenesis of some amino acids (*) results in tetranomial substitution with the wild type, alanine and two additional amino acids (shaded in gray). Whereas shotgun scanning makes use of these tetranomial substitutions in selected positions, binomial mutagenesis either avoids mutagenesis of such residues or applies split-pool synthesis of oligonucleotides encoding mutations.
 tetranomial substitutions of shotgun scanning are employed. Acids, split-pool synthesis of degenerate oligonucleotides or the listed single nucleotide. For combinatorial alanine scanning of other amino acids, (alanine or the wild-type amino acid. Base degeneracies are indicated by R = methyl) lacks unusual backbone dihedral angle preferences; glycine (R = H), for example, would also nullify the role of sidechain functional groups at specific positions can be inferred from alanine mutations. Alanine (sidechain, but could introduce conformational flexibility. The power of alanine-scanning mutagenesis to provide critical biological insight has been demonstrated by the following important early examples: demonstration of its ability to identify functional epitopes. Substitution with alanine (alanine substitution, has been particularly useful for the identification of functional epitopes. Although alanine mutagenesis provides a detailed map of functional epitopes, the method is laborious. Each alanine-substituted protein must be separately constructed, expressed, and sometimes refolded. The loss of the sidechain functionality is then assessed in an in vitro assay of protein activity. In vivo assays can minimize effort spent on protein purification and other steps, but such assays may be available for only a subset of interesting proteins.

Combinatorial libraries of alanine substitutions are an alternative to the laborious method of scanning individual positions in a protein. To apply a combinatorial library technique, two methodological issues must be addressed. Firstly, synthesis of the library requires substitution of alanine and wild-type in specific positions. Secondly, functional proteins must be selected from a library with diversity of up to 10¹¹ alanine-substituted proteins. Approaches to solve these problems draw upon techniques from molecular biology and combinatorial chemistry.

### Table 1

<table>
<thead>
<tr>
<th>Wild-type aa</th>
<th>Replacement aa</th>
<th>DNA codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Ala/Arg/Gly/Pro</td>
<td>(G/C)(G/C)T</td>
</tr>
<tr>
<td>Asn</td>
<td>Ala/Asn/Asp/Thr</td>
<td>(A/G)(A/C)C</td>
</tr>
<tr>
<td>Asp*</td>
<td>Ala/Asp</td>
<td>G(A/C)T</td>
</tr>
<tr>
<td>Cys</td>
<td>Ala/Cys/Gly/Ser</td>
<td>(G/T)(G/C)T</td>
</tr>
<tr>
<td>Glu*</td>
<td>Ala/Glu</td>
<td>G(A/C)A</td>
</tr>
<tr>
<td>Gin</td>
<td>Ala/Gln/Gly/Pro</td>
<td>(G/C)(A/C)T</td>
</tr>
<tr>
<td>Gly*</td>
<td>Ala/Gly</td>
<td>G(G/C)T</td>
</tr>
<tr>
<td>His</td>
<td>Ala/His/Asp/Pro</td>
<td>(G/C)(A/C)T</td>
</tr>
<tr>
<td>Ile</td>
<td>Ala/Ile/Thr/Val</td>
<td>(A/G)(C/T)T</td>
</tr>
<tr>
<td>Leu</td>
<td>Ala/Leu/Pro/Val</td>
<td>(G/C)(C/T)T</td>
</tr>
<tr>
<td>Lys</td>
<td>Ala/Lys/Glu/Thr</td>
<td>(A/G)(A/C)A</td>
</tr>
<tr>
<td>Met</td>
<td>Ala/Met/Thr/Val</td>
<td>(A/G)(C/T)G</td>
</tr>
<tr>
<td>Phe</td>
<td>Ala/Phe/Ser/Val</td>
<td>(G/T)(C/T)T</td>
</tr>
<tr>
<td>Pro*</td>
<td>Ala/Pro</td>
<td>(G/C)CA</td>
</tr>
<tr>
<td>Ser*</td>
<td>Ala/Ser</td>
<td>(G/T)CC</td>
</tr>
<tr>
<td>Thr*</td>
<td>Ala/Thr</td>
<td>(A/G)CT</td>
</tr>
<tr>
<td>Trp</td>
<td>Ala/Trp/Gly/Ser</td>
<td>(G/T)(G/C)G</td>
</tr>
<tr>
<td>Tyr</td>
<td>Ala/Tyr/Asp/Ser</td>
<td>(G/T)(A/C)T</td>
</tr>
<tr>
<td>Val*</td>
<td>Ala/Val</td>
<td>G(C/T)T</td>
</tr>
</tbody>
</table>

*For each amino acid (aa), a degenerate codon is used to substitute alanine or the wild-type amino acid. Base degeneracies are indicated by parentheses (e.g. (G/C) represents a 1:1 ratio of bases G and C). For seven amino acids, binomial substitution is accessible by exchanging a single nucleotide. For combinatorial alanine scanning of other amino acids, split-pool synthesis of degenerate oligonucleotides or the listed tetranomial substitutions of shotgun scanning are employed."

The functional epitope reveals how a protein works. Understanding protein function also requires identification of residues that position the sidechains in direct contact (‘second sphere residues’) [12]. In addition, biophysicists are also interested in residues critical to protein folding and stability. Combinatorial alanine-scanning (Figure 1) is ideally suited to identify all such important residues.

### Alanine mutagenesis

Alanine-scanning mutagenesis, a method of systematic alanine substitution, has been particularly useful for the identification of functional epitopes. Substitution with alanine removes all sidechain atoms past the β-carbon. Thus, the role of sidechain functional groups at specific positions can be inferred from alanine mutations. Alanine (sidechain R = methyl) lacks unusual backbone dihedral angle preferences; glycine (R = H), for example, would also nullify the sidechain, but could introduce conformational flexibility into the protein backbone. The power of alanine-scanning mutagenesis to provide critical biological insights has been demonstrated by the following important early examples: human growth hormone (hGH) binding to hGH-binding protein (hGHbp) [13]; CD4-binding to HIV–gp120 [14]; the enzymatic activity of kinases [15]; and lysozyme stability [16], amongst many others.

Combinatorial libraries of alanine substitutions are an alternative to conventional alanine-scanning mutagenesis, several research groups have described methods to access multiple alanine substitutions. Through a single round of site-specific oligonucleotide-directed mutagenesis, binomial substitutions of either alanine or a wild-type amino acid are readily accessible by conventional oligonucleotide synthesis for seven amino acids (labeled with an asterisk in Table 1). For these seven amino acids, altering a single encoding nucleotide can result in a codon for alanine. For example, the codon TCC encodes the amino acid serine; replacing T with G in the first position results in a codon for alanine (GCC). Between oligonucleotide synthesis, addition of a 1:1 ratio of the T and G phosphoramidites in the first position of the codon will result in a 1:1 ratio of serine and alanine in the translated protein library. Libraries with alanine substitutions in multiple positions can be encoded by degenerate oligonucleotides with mutations in multiple positions.

### Combinatorial, site-specific mutagenesis

**Binomial mutagenesis**

As an alternative to conventional alanine-scanning mutagenesis, several research groups have described methods to access multiple alanine substitutions. Through a single round of site-specific oligonucleotide-directed mutagenesis, binomial substitutions of either alanine or a wild-type amino acid are readily accessible by conventional oligonucleotide synthesis for seven amino acids (labeled with an asterisk in Table 1). For these seven amino acids, altering a single encoding nucleotide can result in a codon for alanine. For example, the codon TCC encodes the amino acid serine; replacing with G in the first position results in a codon for alanine (GCC). Between oligonucleotide synthesis, addition of a 1:1 ratio of the T and G phosphoramidites in the first position of the codon will result in a 1:1 ratio of serine and alanine in the translated protein library. Libraries with alanine substitutions in multiple positions can be encoded by degenerate oligonucleotides with mutations in multiple positions.

Combinatorial binomial mutagenesis has been used skillfully by Gregoret and Sauer [17] to analyze the consequences of multiple alanine substitutions upon the stability and function of the DNA-binding protein λ repressor. 11 positions of the helix-turn-helix of λ repressor, a critical region for protein functionality, were mutated to alanine or wild type (Figure 2a). Approximately 25% of the alanine-substituted proteins retained activity, an indication of the robust information encoding protein folding. The combinatorial format of alanine substitutions also tested a long-debated question — are the energetic contributions of individual amino acids additive? Binomial mutagenesis of λ repressor, shotgun scanning (described below), and work by others [18] has demonstrated conclusively that, with a few exceptions, the majority of individual sidechains contribute binding energy to the receptor–ligand interaction in an...
Combinatorial chemistry

Vernet and co-workers \[21,22\] reported the split-pool synthesis of degenerate oligonucleotides for combinatorial alanine mutagenesis. Split-pool synthesis of oligonucleotides can be used to synthesize an alanine codon in one pool and the wild-type codon in a different pool. Combinatorial alanine-scanning with mutagenesis by split-pool synthesized oligonucleotides was used to investigate the interface between the heavy and light variable domains of an antibody (Figure 2b). The functional requirement for conservation of wild-type sidechains bordering the antigen-binding site was ascribed to second-sphere effects by Altschuh, Vernet and co-workers \[23\]. These results demonstrate the usefulness of combinatorial alanine-scanning for identification of residues contributing indirectly to protein function.

The resilience of protein folding despite extensive alanine substitutions has been demonstrated by the experiments described above, multiple alanine substitutions of T4 lysozyme \[24\], Arc repressor \[25\] and other systems (reviewed in \[26\]). This research set the stage for combinatorial alanine mutagenesis techniques such as shotgun scanning. Shotgun scanning can apply combinatorial alanine mutagenesis to more than 20 positions, resulting in exceptionally diverse protein libraries.

**Shotgun scanning**

Shotgun scanning (compared with binomial mutagenesis in Table 2) implements a simplified format for combinatorial alanine-scanning. Libraries of alanine-substituted proteins are displayed on the surface of filamentous phage particles for \textit{in vitro} selections (Figure 3). By displaying libraries of alanine-substituted proteins on the surface of filamentous phage particles, successive rounds of a binding selection are used to enrich residues contributing binding energy to the receptor–ligand interaction. Conventional, automated DNA synthesis is used for the synthesis of mutagenic oligonucleotides encoding the shotgun scanning libraries.

Phage display, first described in 1985 by George Smith \[27\], simplifies the construction and screening of combinatorial libraries of alanine-substituted proteins in at least three important ways (reviewed in \[28,29\]). First, large libraries of proteins (>10\(^{10}\) unique clones) are readily accessible, with each unique protein variant fused to the surface of a different phage particle. After selection for displayed proteins that bind to a receptor, phage can be amplified in
Combinatorial alanine-scanning
Morrison and Weiss 305

an *Escherichia coli* host. Lastly, the phage particles encapsulate DNA encoding the displayed protein; thus, standard DNA sequencing can be used to identify selected proteins. Shotgun-scanning libraries substitute either the wild-type amino acid, alanine or up to two other amino acid sidechains in each specifically mutated position (Table 1) [30••].

Shotgun scanning differs from previous combinatorial alanine-mutagenesis techniques by application of conventional oligonucleotide synthesis and substitution of some amino acids with four possible sidechains. A key simplifying assumption is made during analysis of the selected clones. Analysis of the energetic contribution by specific sidechains to receptor–ligand binding focuses entirely on the distribution of alanine or wild-type in each substituted position. With this simplification, combinatorial alanine mutagenesis becomes possible using standard oligonucleotide synthesis. However, secondary analysis of non-wild-type, non-alanine substitutions in each position is also possible. Statistical analysis to parse information about protein functionality from tetranomial mutagenesis has been described by Tidor, Sauer and co-workers [31]. Strong selection for non-alanine and non-wild-type amino acids could reveal unexpected interactions, such as mutations that confer improved affinity to the ligand.

The first account of shotgun scanning tested this focus on the distribution of wild type or alanine in each position. The 19 residues of hGH comprising the high-affinity binding site for hGHbp were shotgun-scanned in a single library (Figure 2c). After multiple rounds of selection and amplification in an *E. coli* host, individual hGHbp-binding phage were identified by a high-throughput binding assay and subjected to DNA sequencing. Focussing entirely upon the distribution of alanine or wild type in each scanned position revealed specific positions that were highly conserved as the wild-type amino acid, whereas other positions demonstrated a roughly even distribution of alanine or wild type, hGH was chosen for this initial experiment to allow comparison between shotgun scanning and conventional alanine-scanning mutagenesis.

### Table 2: Comparison of binomial mutagenesis and shotgun scanning.

<table>
<thead>
<tr>
<th>Mutagenesis method</th>
<th>Oligonucleotide-directed</th>
<th>Library format (reported to date)</th>
<th>Screening (reported to date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial mutagenesis</td>
<td>2: Wild-type and alanine</td>
<td><em>E. coli</em></td>
<td>In vivo</td>
</tr>
<tr>
<td>Shotgun scanning</td>
<td>2 to 4: Wild-type or alanine and 2 additional in some positions (see Table 1)</td>
<td>Phage display</td>
<td>In vitro</td>
</tr>
</tbody>
</table>

Data from hGH shotgun-scanning compare favorably with alanine-scanning mutagenesis. A caveat to shotgun scanning is that the library selection by binding to a receptor requires equilibrium conditions for results to be comparable to values obtained from individual alanine substitutions. ∆∆G values compare the binding energy for the wild-type receptor–ligand interaction with those of the alanine-substituted protein, and, thus, quantify the contribution to binding energy made by atoms past the β-carbon. By assuming that the shotgun scanning selection occurred during equilibrium binding conditions, pseudo-∆∆G values from shotgun scanning were derived from the distribution of alanine and wild-type amino acids in each position. The shotgun scanning pseudo-∆∆G values confirmed that hGH binding to hGHbp is mediated by a compact hot spot of just seven residues out of the 19 residues in contact with the receptor. Shotgun scanning offered the convenience of a single round of mutagenesis in multiple positions with the rapid purification and assay of phage-displayed alanine-substituted proteins [30••].

### Future directions

Recent interest in combinatorial alanine-mutagenesis (e.g., shotgun scanning) is fueled by breakthroughs in DNA

---

**Figure 3**

Shotgun scanning. (a) Libraries of proteins with alanine and the wild-type amino acid are constructed, using oligonucleotide-based, site-directed mutagenesis. (b) The pooled library is applied to a solid support-bound receptor and non-binding phage are washed away. (c) Through successive rounds of a binding selection and amplification in an *E. coli* host, wild-type sidechains are enriched in positions with energetically favorable contacts. Conversely, no enrichment for the wild-type sidechain was observed for positions failing to contribute binding energy to the hGHbp–hGH interaction.

---

**Table 2**

Comparison of binomial mutagenesis and shotgun scanning.

<table>
<thead>
<tr>
<th>Mutagenesis method</th>
<th>Oligonucleotide-directed</th>
<th>Library format (reported to date)</th>
<th>Screening (reported to date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial mutagenesis</td>
<td>2: Wild-type and alanine</td>
<td><em>E. coli</em></td>
<td>In vivo</td>
</tr>
<tr>
<td>Shotgun scanning</td>
<td>2 to 4: Wild-type or alanine and 2 additional in some positions (see Table 1)</td>
<td>Phage display</td>
<td>In vitro</td>
</tr>
</tbody>
</table>
sequencing, as statistical analysis of the wild-type to alanine ratio requires extensive DNA sequencing. With continuing improvements to DNA sequencing technology, more examples of shotgun scanning and related techniques will emerge. Combinatorial alaniine-mutagenesis provides faster analysis of receptor–ligand interactions than conventional alaniine-scanning and offers analysis within the context of multiple mutations. Shotgun scanning makes this technique available in a format applicable to a wide range of receptor–ligand interactions. With shot- gun scanning, combinatorial alaniine mutagenesis requires neither an in vivo selection, nor specialized synthesis of the mutation-encoding DNA. In addition, the phage-display format of shotgun scanning offers a rapid selection and assay of the alaniine mutations. Beyond simple convenience, shotgun scanning could extend analysis of protein function with unprecedented scope and detail.

In the past, systematic mutagenesis of many residues in a protein has required a tour de force effort ([32–34]; reviewed in [35]). With shotgun scanning and other combinatorial techniques becoming routinely available, whole-protein mutagenesis experiments can be planned. For example, a 120-residue protein can be divided into six shotgun scanning libraries, each with 20 residues substituted as alanine or wild type. While alaniine mutagenesis highlights the contribution made by the functional group of each sidechain, substitution with other amino acids can offer different insight into protein function. Shotgun scanning with the α-helix breaker proline, for example, could probe the functional contribution of each position in an α helix. While the examples of combinatorial alaniine mutagenesis discussed here have emphasized analysis of protein stability and non-covalent binding interactions, such techniques could be used to explore enzymatic catalysis, multi-subunit assembly and other aspects of protein function.

Conclusions

With the recently completed DNA sequencing phase of the human genome project and the expected explosion of protein structures from structural genomics efforts, methods are needed for rapid, yet detailed, analysis of receptor–ligand interactions. Combinatorial alaniine-scanning combines the expedience of combinatorial libraries with the insight of site-directed scanning mutagenesis. The recently reported shotgun scanning format for combinatorial alaniine-scanning can be used to map functional epitopes and rapidly assign binding energy to individual functional groups at receptor–ligand interfaces. Intensive analysis of receptor–ligand interactions by combinatorial alaniine-scanning will elucidate principles of molecular recognition, protein folding, and the relationship of protein function to protein structure.

Acknowledgement

The authors thank the University of California at Irvine, School of Physical Sciences for support.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
•• of outstanding interest


29. Sidhu SS: Phage display in pharmaceutical biotechnology. Curr Opin Biotechnol 2000, 11:810-816. The author reviews many recent technological improvements to phage display. These advances broaden the applicability of phage display, which in turn extends the range of receptor–ligand interactions amenable to shotgun scanning and other combinatorial alanine-scanning techniques.

30. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS: Rapid mapping of functional protein epitopes by combinatorial alanine-scanning. Proc Natl Acad Sci USA 2000, 97:8950-8954. This paper reports shotgun scanning, combinatorial alanine-scanning in a phage display format. Data from shotgun scanning of the hGH/hGHbp interaction compared well with analysis by conventional alanine-scanning. Shotgun scanning requires only that the proteins of interest be amenable to phage display and in vitro screening. Therefore, this technique should allow rapid functional epitope mapping of a broad range of receptor–ligand interactions.


