Using Peptoid Libraries [Oligo N-Substituted Glycines] for Drug Discovery

Reyna J. Simon, Eric J. Martin, Susan M. Miller[a], Ronald N. Zuckermann, Jeffrey M. Blaney, and Walter H. Moos

Chiron Corporation, Emeryville, CA 94608
[a]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

I. Introduction

In pharmaceutical research, screening of compound libraries has been a fruitful method for the discovery of lead compounds. Recent advances in the preparation of drug candidates via automated and combinatorial methods have allowed unprecedented numbers of compounds to be generated and screened—the compound libraries of the nineties. This blossoming field of research has been dubbed "molecular diversity" in recognition of the breadth of compounds currently being tested.

However, molecular diversity is not a new concept. As early as the 1940s, pharmaceutical companies realized the therapeutic efficacy and commercial potential of using natural products as pharmaceuticals (1). These compounds are typically isolated from bacterial fermentation broths, fungi, plants, insects and marine organisms. Penicillin is one of the most recognizable drugs from that era. As exemplified by taxol, natural products continue to yield potent pharmaceuticals. Perhaps by virtue of their natural origin, these compounds are quite diverse, with chemical structures often containing many stereocenters. Due to the complicated nature of the chemicals, identification of lead compounds from these sources is labor intensive, and lead development is often a synthetically challenging and costly task.

Another source of molecular diversity stems from pharmaceutical companies amassing large numbers of synthetic targets and intermediates from various research programs. Containing up to a million compounds, with new compounds being added at perhaps several thousand per year, these individual company libraries constitute an important proprietary resource. Screened against the biological assay *du jour*, the libraries have provided significant chemical diversity that only comes from many person-years of work. New companies are clearly at a disadvantage in this respect, having neither the number nor diversity of compounds in their stocks.

In the late 1980s, technology in solid phase compound synthesis of both peptides and oligonucleotides advanced to the point where new players could actively compete in library generation and screening. For peptides, parallel and mixing synthetic strategies have been exploited to generate libraries of greater than tens of thousands of compounds *per experiment*. The combinatorial relationship between the length of an oligomer and the number of variants at each position has made the synthesis of large libraries possible. An oligomer of length n with x variants at each position will give x^n compounds. With relative ease, libraries of peptides and oligonucleotides have rapidly been synthesized and used in drug research. The drawback is that, unlike the natural product or synthetic compound libraries traditionally used, these libraries contain only natural biopolymers: peptides and oligonucleotides. The metabolic instability of these compounds owing to proteolysis or nuclease activity and their poor absorption characteristics often render them poor drug candidates.

TECHNIQUES IN PROTEIN CHEMISTRY V

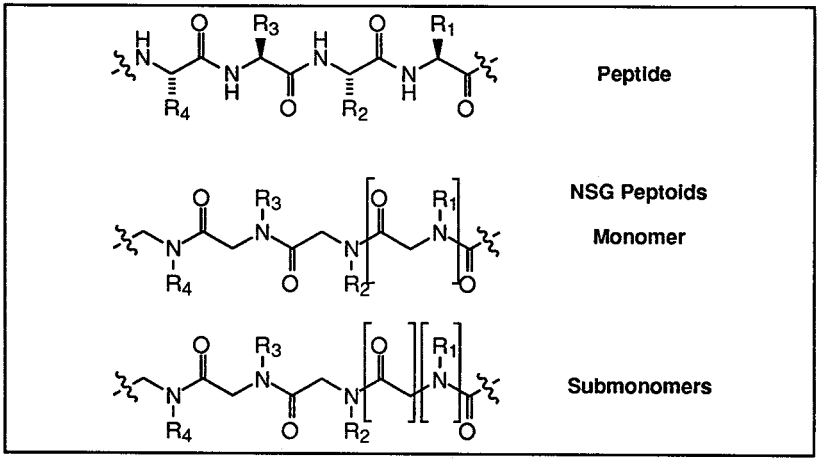

Figure 1. Top frame shows a representative peptide. Bottom frame shows two NSG peptoids, with the NSG monomer and submonomers outlined.

The goal of our research is to create a chemical alternative to natural biopolymers that would be pharmaceutically relevant, but which also retains the features of solid phase chemistry, namely: modular stepwise assembly from synthetically accessible monomers, capacity for automation, and rapid generation of diverse libraries. The present paper focuses on "peptoids" as one solution to this problem: the design, synthesis, and use of diverse peptoid libraries. The name peptoid was coined in 1982 (2) and refined in 1988 (3) to mean a credible drug candidate that mimics or blocks the action(s) of an endogenous peptide. This definition is broad and encompasses many possible chemical entities. We have applied the term to a new class of compounds that consist of oligomers of N-substituted glycines (NSG peptoids) (4).

Table I. Table of Amino Acid Like NSG Peptoid Monomers Prepared (4)

Sidechain (R)	P[a]	Designator	Method[b]
CH_3		Nala	*c*
$CH_2CH_2CH_2N{=}C(NH_2)_2$	Pmc	Narg	A
CH_2CO_2H	t-Bu	Nasp	A
$CH_2CH_2CO_2H$	t-Bu	Nglu	A
$CH_2CH_2CONH_2$		Ngln	C
CH_2CH_2-(4-imidazolyl)	Trt	Nhhis	A[d]
$CH_2CH(CH_2)_2$		Nleu	B
$CH_2CH_2NH_2$	Boc	Naeg	B
$CH_2C_6H_5$		Nphe	*c*
CH_2CH_2OH	t-Bu	Nhser	A
CH_2CH_2-(3-indolyl)	Boc	Nhtrp	A
CH_2CH_2-p-C_6H_4OH	t-Bu	Nhtyr	A

[a] Sidechain protecting group introduced prior to assembly of N-substituted glycine. [b] See Figure 2. [c] Commercially available. [d] $NaBH_3CN$ used in place of H_2/Pd.

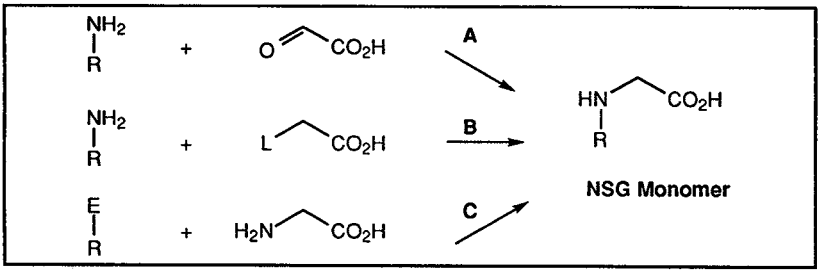

Figure 2. General routes to NSG monomers. R is protected prior to assembly of NSG. A: H_2, Pd/C or $NaBH_3CN$; B: Alkylation with excess RNH_2; C: Michael conditions to form Ngln where E-R is acrylamide.

II. The Modular Approach to Drug Discovery

Peptides are repeating oligomers that can be schematically represented as a set of building blocks linked by a chemical linking unit, *e.g.*, the amino acids are linked via amide bonds. The sidechains of the amino acids are chemical branches off of the polyamide backbone. In our modular approach, there may be many varieties of backbones as well as many types of sidechains. A simple extension of this concept led to the development of N-substituted glycine oligomers, or NSG peptoids. The monomer unit is an N-substituted glycine. As shown in Figure 1, the relationship to peptides is obvious; the sidechains of the amino acids are moved from the alpha-carbon to the nitrogen. In recent years, other groups have used similar concepts in the designs of new types of oligomers, namely peptide nucleic acids, vinylogous peptides, etc. (5)

A series of compounds for solid phase synthesis have been prepared with functional groups similar to or identical to the natural amino acids (Table I) with the routes for their syntheses shown in Figure 2 (4). In general, either reductive amination, route A, or alkylation, route B, provide the necessary monomers. These are then incorporated into a growing peptide/peptoid chain using the activating agents, BOP or PyBroP (6) (Figure 3A), under conditions similar to the Merrifield approach to peptide synthesis. Yields and characteristics of the products are analogous to those of peptides. Oligomeric NSG peptoids and peptide/peptoid chimeras have been prepared with activity in several different bioassays (4).

Synthesis can also be accomplished using a different scheme (7). The NSG oligomer can be thought of as an alternating copolymer of acetate/amine units rather than an oligomer of NSG units (Figure 1 and 3B). The NSG monomers are subdivided into two units. These *submonomers* can be sequentially reacted to give a growing peptoid chain identical to that synthesized by traditional Merrifield methods. Using this route, small oligomers can be routinely prepared, and a 25mer has been prepared in good yield and purity (7).

There are several advantages to the submonomer route. The building blocks are primary amines, which are commercially available and functionally diverse. With this synthesis scheme, many different amines can be used, giving final products that may resemble peptides *solely* by virtue of the polyamide backbone. Thus, the synthesis proceeds from inexpensive and readily available starting materials, and allows for the creation of diverse chemical libraries.

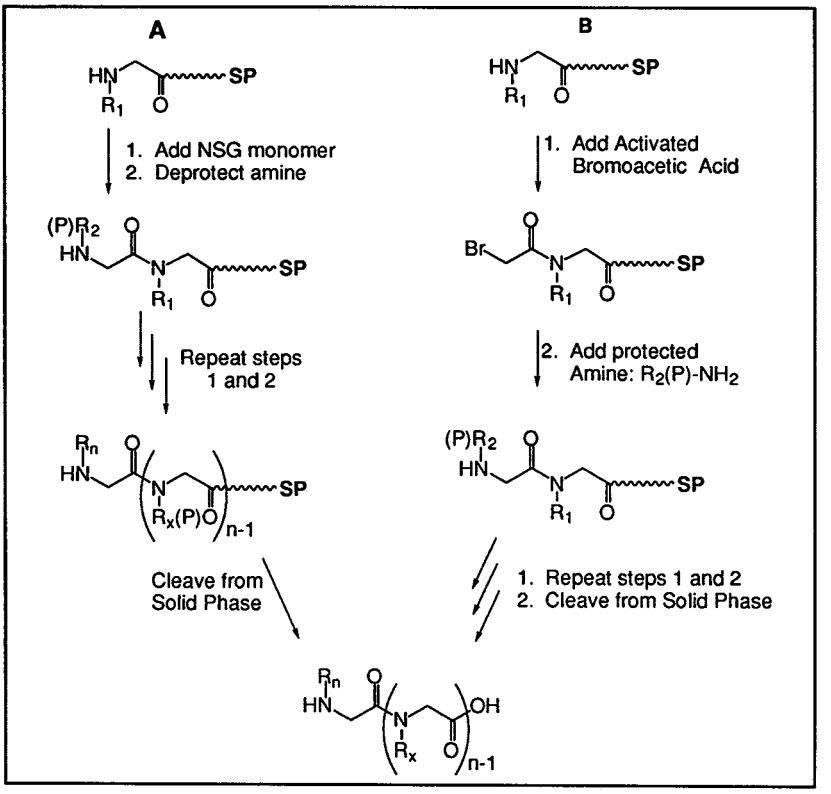

Figure 3. A: Solid phase peptoid synthesis under Merrifield conditions; R(P) is a protected sidechain. B: Submonomer solid phase peptoid synthesis.

III. Optimization of Combinatorial Libraries

Given that the submonomer synthesis allows any of over 1000 readily available primary amines to be used as peptoid side chains, over 10^{12} tetramers can be made. Even if combinatorial libraries were synthesized as mixtures and screened in automated receptor binding assays at rates exceeding 10,000 compounds per week, millions of years would be required to screen the entire pool. Thus, a strategy is needed *a priori* to select specifically which diverse libraries should be made to maximize the chances of finding a lead during initial screening, and which focused libraries should be generated *a posteriori* to optimize a lead that is already known.

Efficient broad screening uses combinatorial oligomer libraries made from all combinations of a set of monomers that are maximally dissimilar to each other. This minimizes redundancy in the test set. Highly focused screening uses oligomer libraries where each position is sampled from a set of monomers most similar to the sidechain at that position in a lead peptoid. These and other intermediate strategies are all based on the ability to measure similarity and dissimilarity between potential monomers and to select sets that maximize or minimize this resemblance. This can be accomplished by computing a variety of

properties for each sidechain such that the distance between property vectors reflects the similarity between the monomers. Designs are then chosen to maximize or minimize these distances. Of course, a computed set of properties may not account for every biologically important feature. The quantitative method, therefore, must also provide a mechanism to incorporate the intuition of experienced chemists and biologists.

In this work, the calculated properties include lipophilicity, shape and branching, functional group descriptors, and "atom layer properties", each of which is described below. Lipophilicity is calculated for the neutral amines using the CLOGP (8,9) and HINT (10) programs. These programs fail for many compounds because they are empirical fragment-based methods. Where no other method succeeded, the remaining properties are estimated by comparison to experimental values for analogous compounds in the Pomona92 database (11).

Shape and branching are characterized by a collection of topological indexes (12) computed with the Molconn-X program (13). These are reduced to 5 principle components that reproduced 80% of the variance in these 73 properties. Although the exact chemical significance of any one of these variables may be obscure, compounds with a similar set of topological property vectors are quite similar, whereas compounds with very different sets of shape descriptors are quite distinct.

Substructure search keys from chemical databases are a rich source of local chemical functional group information. The Daylight substructure "fingerprints" use a hashing algorithm to assign all substructures up to 7 bonds long to binary indicator variables in a 2048 bit string (14). How is a 2048 dimensional binary description reduced to a small number of continuous variables for use in experimental design? The solution is to use the fingerprints to compute the Tanimoto similarity coefficient (15) between each monomer pair. Multidimensional scaling (MDS) is then used to create a low dimensional "map" that reflects these "distances." MDS is a statistical technique akin to the embedding and refinement steps of distance geometry. For each member of a set, it calculates Cartesian coordinates in some low dimensional space that best reproduce a table of distances. In distance geometry, these would be atom position coordinates that best reproduce interatomic distances. In our problem, these are latent chemical functionality variables that best reproduce intermonomer dissimilarities. MDS reduces the 2048 bits of information to 5 latent "functional group" variables that reproduce the original Tanimoto similarities with a relative standard deviation of only 10%.

The above properties are largely insensitive to the point where the sidechain attaches to the peptoid backbone, and do not account for many specific chemical interactions that are believed to be important in pharmacophore recognition. "Atom layer" properties were developed to address these concerns. An "atom layer" consists of the set of all atoms a given number of bonds away from the backbone. Using a simple empirical expert system developed using the Daylight tool kit (16), each non-hydrogen atom in a monomer is characterized by its radius and whether it is an acid, a base, an H-bond donor, an H-bond acceptor and/or an aromatic moiety. Each of the six atom properties are summed for the atoms in each layer (up to 15 layers deep). Thus, each monomer is characterized by a table of 6 properties by 15 atom layers. The tables are compared by a generalization of Tanimoto similarity. MDS is applied to the resulting distance matrix exactly as in the substructure fingerprints above. Five dimensions again seem to be a useful number.

The final task is to choose small subsets of molecules with maximal similarity or dissimilarity. All of the properties described above are combined into a single property table in which each monomer from the pool of 1000 is represented by a vector of the 16 properties described above: 5 finger print MDS dimensions, 5 atom layer MDS dimensions, 5 topological index principle components and log P. Sets of monomers similar to a lead are chosen simply by

rank ordering every member of the pool by the Euclidean distance from the reference target.

Finding dissimilar sets is more difficult. In particular, we wanted to specify a small collection of interesting monomers, *e.g.*, monomers with known pharmacophores for the receptor of choice, then fill out the rest of the set with a few monomers from the full pool of 1000 that are diverse both from the initial set and from each other. This is accomplished with D-optimal design techniques. Roughly speaking, in order to determine accurate parameter estimates to a response surface, the D-optimal design algorithm chooses subsets of points from a larger pool that are well spread out and largely orthogonal in property space, *i.e.*, they are diverse. The method is flexible in that any number of monomers can be specified for initial inclusion in the set. The optimizer will then select the additional points that best fill out the set to a specified size. This allows the essential marriage of chemical intuition and quantitative statistical design.

These properties have now been calculated for the commercially available amines that are suitable for submonomer synthesis, *i.e.*, the amines that are inexpensive, readily available, and react well in model syntheses. This method is routinely used in designing diverse mixtures for initial screening, or focused mixtures similar to a given active peptoid for lead optimization.

IV. Pharmacokinetic Considerations

Four properties of potential therapeutics which often need to be maximized for a drug to be therapeutically useful are: absorption, distribution, metabolism and excretion. Many peptide or peptide mimetic drug candidates have not been developed due to poor absorption or metabolic instability (17). For certain therapies, NSG peptoids may prove to be superior to peptides in all of these categories.

The *in vivo* lifetime of a bioactive peptide can be significantly extended if the peptide is modified at the amide bond most susceptible to proteolysis. N-alkylation is often sufficient to prevent degradation. By virtue of the complete N-substitution of the amide backbone in NSG peptoids, they were expected to be resistant to proteases, thereby making them good candidates for extended *in vivo* lifetimes. We sought to systematically investigate this problem with a series of peptides and peptoids and several common proteases.

A series of oligomers were prepared in several varieties: the L peptide, D peptide, NSG peptoid, and reverse sequence NSG peptoid (18). These are listed in Table II. The sequences of the all L peptides were chosen to reflect likely proteolysis candidates for the corresponding proteases: chymotrypsin, elastase,

Table II. Sequences of Proteolysis Peptides

Sequence[a]	Protease
Ac-Ala-Ala-Ala-Leu-Phe-Arg-NH_2	Elastase
Ac-Ala-Leu-Phe-Ala-Leu-Arg-NH_2	Chymotrypsin
Ac-Phe-Ala-Arg-Ala-Arg-Asp-NH_2	Trypsin
Ac-Ala-Phe-Glu-Leu-Ala-Ala-NH_2	Papain
Cbz-Ala-His-Phe-Phe-Arg-Leu-NH_2	Pepsin[b]

[a] Sequences were prepared as the L, D, N, and reverse N hexamers. In all cases except for His, the NSG sidechain was identical to the amino acid sidechain. For His, the Nhhis (extra CH_2) was used. AAA obtained for peptides. MS obtained for peptoids. [b] Pepsin reverse N hexamer not prepared.

pepsin, papain, and trypsin. Aliquots of the proteolysis mixtures were assayed with 4-fluoro-7-nitro-benzofurazan which reacts with free amines to give a fluorescent product (19). N-capped oligomers were used as substrates. Therefore only proteolyzed oligomers fluoresced under these conditions. The all L isomers were proteolyzed by their respective enzymes, whereas the D peptide and the NSG peptoids were untouched.

We are currently addressing the bioavailability of NSG peptoids. Based on a recent report, Conradi and coworkers have examined the membrane permeability of a series of N-substituted peptides (20). Membrane permeability is one measure of the ability of a compound to be absorbed in the GI tract, thus contributing to its bioavailability. The transport of model peptides and N-substituted peptides was measured across confluent monolayers of Caco-2 cells as a model of the intestinal mucosa. The results are consistent with a hydrogen bond model of transport, where hydrogen bond donors (*i.e.*, amide hydrogens) need to be desolvated before membrane permeation can occur. In NSG peptoids, there are *no* backbone hydrogen bonds. By analogy, these molecules would traverse a membrane more readily than peptides, suggesting improved bioavailability over peptides.

In conclusion, NSG peptoid libraries are synthesized from readily available starting materials, allowing the rapid generation of diverse collections of compounds. With their improved pharmacological properties, these libraries should provide novel entries into the drug discovery field with better prospects for discovering orally bioavailable compounds.

References

1. Moos, WH, Green, G, Pavia, MR (1993). Annu. Rep. Med. Chem. **28**, in press.
2. Farmer, PS, Ariens, EJ (1983). Trends Pharmacol. Sci. **3**, 362-5.
3. Horwell, DC (1988). Spec. R. Soc. Chem. **65**, 62-72.
4. Simon, RJ, *et al.* (1992). Proc. Natl. Acad. Sci. USA **89**, 9367-71.
5. This subject has been recently reviewed. See Zuckermann, RN (1993). Current Biology **3**, 580-4.
6. PyBOP is benzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate. PyBroP is bromotris(pyrrolidino)-phosphonium hexafluorophosphate.
7. Zuckermann, RN, *et al.* (1992). J. Am. Chem. Soc. **114**, 10646-7.
8. (1993) *In* CLOGP v. 431, Daylight Chemical Information Systems, Irvine, CA.
9. Leo, A (1993). Chem. Rev. **93**, 1281.
10. Kellogg, GE, Joshi, GS, and Abraham, DJ (1992). Med. Chem. Res. **1**, 444-53.
11. (1992) Pomona 92 database, Daylight Chemical Information Systems, Irvine, CA.
12. Hall, LH, and Kier, LB (1991). *In* "Reviews in Computational Chemistry" (Lipkowitz and Boyd ed.), Vol. 2, p. 367-421, VCH Publishers, Inc.
13. Molconn-X ver. 1.0, Hall Associates consulting, Quincy, MA.(1991).
14. James, CA, and Weininger, D (1992). *In* "Daylight 4.3 Theory Manual," Daylight Chemical Information Systems, Inc., Irvine, CA.
15. Willett, P (1987). *In* "Similarity and Clustering in Chemical Information Systems," p 54. John Wiley & Sons, New York.
16. James, CA, *et al.* (1992). *In* "Daylight Toolkit Programmer's Guide," Ver. 4.31, Daylight Chemical Information Systems, Inc., Irvine, CA.
17. Fauchere, J, and Thuriear, C (1992). Adv. in Drug Res. **23**, 127-59.
18. The reverse peptoids (reverse sequence) were also prepared because of the ambiguity of the alignment of the oligomers when matching up the sidechains of peptoids and peptides. See reference 4.
19. Toyooka, T, *et al.* (1983). Anal. Chim. Acta **149**, 305-12; and Miyano, H, *et al.* (1985) ibid. **170**, 81-7.
20. Conradi, J, *et al.* (1992). Pharm. Res. **9**, 435-8.