mustard from a structurally diverse range of parent molecules. This enables the adjustment of the parent compounds redox potential to match that of the hypoxic cancer cell without interfering with the DNA crosslinking ability of the DNA mustard.

5 Anon (2000) Nitrogen mustard derivatives of (1,4-benzoquinonyl)alkanoic acids as hypoxia-sensitive antitumour agents. *Exp. Opin. Ther. Patents* 10, 507–511

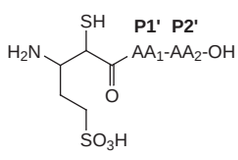
Andrew Lloyd

Combinatorial chemistry

Subsite preferences in aminopeptidase A

Aminopeptidase A (APA) or EC 3.4.11.7, is a membrane-bound zinc metallopeptidase that specifically cleaves acidic N-terminal amino acids from peptide substrates. The enzyme has significant homology with aminopeptidase N (APN), another peptidase that cleaves hydrophobic and basic N-terminal residues from peptides. To gain a better understanding of the physiological function of APA in brain and peripheral tissues it is necessary to identify efficient and selective inhibitors. A combinatorial approach has been used to investigate the subsite preferences of APA (Ref. 1).

Amastatin is generally used as an APA inhibitor but this compound is not selective and, in particular, has affinity for APN. This study investigated the combinatorial synthesis of thiol-containing compounds of the structure (**i**). It was found that the introduction of a



(i)

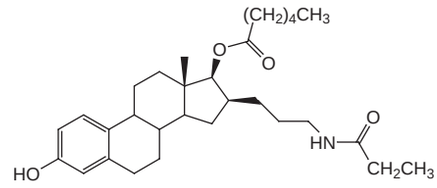
sulphonate into the P1 position, a hydrophobic group into P1′, and a (3*R*)-carboxyproline in P2′ gave rise to highly selective and efficient inhibitors of APA.

1 David, C. *et al.* (1999) Investigation of subsite preferences in aminopeptidase A (EC 3.4.11.7) led to the design of the first highly potent and selective inhibitors of this enzyme. *J. Med. Chem.* 42, 5197–5211

Solid-phase synthesis of phenolic steroids

Both 17β-hydroxysteroid dehydrogenase and the oestrogen receptor can bind oestradiol as a natural ligand/substrate and appear to play key roles in oestrogen-sensitive diseases such as breast and endometrium cancers. These two targets can be blocked by two different drugs or the same drug, offering a novel therapy for these cancers. A recent paper describes a method for the solid-phase combinatorial synthesis of phenolic steroids with relevance for these protein targets[2].

A survey of possible solid-phase linker methods for the steroid phenol group revealed that a photolabile *o*-nitrobenzyl ether linker was most effective in generating products (e.g. **ii**) in excellent yields and purities. This study



(ii)

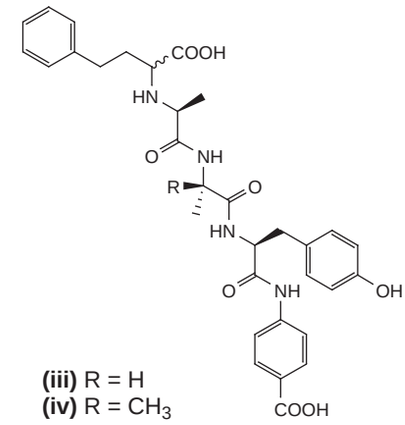
of library methods has set the scene for the generation of libraries of oestradiol-related compounds that could be tested for affinity for inhibition of the oestradiol binding proteins.

2 Tremblay, M.R. and Poirier, D. (2000) Solid-phase synthesis of phenolic steroids: from optimization studies to a convenient procedure for combinatorial synthesis of biologically relevant estradiol derivatives. *J. Comb. Chem.* 2, 48–65

Inhibitors of EC 3.4.24.15

In vitro, the neutral metalloendopeptidase EC 3.4.24.15 hydrolyses several biologically active peptides including bradykinin and neurotensin. Although it might have a physiological role in brain and endocrine function, further investigations of its function have been restricted by the lack of a stable potent inhibitor. To date, the use of the most frequently used inhibitor, *N*-[1-(*R,S*) carboxy-3-phenylpropyl]-Ala-Ala-Tyr-*p*-aminobenzoate (cFP, **iii**), has been limited by rapid hydrolysis of the Ala-Tyr bond. Therefore, a recent study has investigated the design and solid-phase synthesis of novel stable inhibitors of EC 3.4.24.15 (Ref. 3).

cFP was used as a template for the solid-phase preparation of compounds in which the scissile bond has been replaced by groups that are more stable to hydrolysis. The compound in which the Ala-Tyr amide bond had been reduced to an aminomethyl group had a reduced affinity for the enzyme by some thousand-fold. However, replacement of the Ala residue with aminoisobutyric acid produced compound (**iv**) with a K_i of 23 nM. Furthermore, this compound is stable to hydrolysis and does not inhibit angiotensin converting enzyme, or other related thermolysin-like or neutral endopeptidases. The compounds therefore provide a valuable tool for the further investigation of the physiological function of EC 3.4.24.15.



(iii) R = H
(iv) R = CH_3

1359-6446/00/$ – see front matter PII: S1359-6446(00)01505-1

3 Shrimpton, C.N. *et al.* (2000) Development and characterization of novel potent and stable inhibitors of endopeptidase EC 3.4.24.15. *Biochem. J.* 345, 351–356

Nick Terrett
Discovery Chemistry
Pfizer Central Research
Sandwich, Kent, UK
fax: +44 1304 655419
e-mail: nick_terrett@sandwich.pfizer.com

Bioinformatics

Multiple alignment – an essential bioinformatic tool

Multiple alignment is an important and essential tool for bioinformatics research[1,2]. Multiple alignments are alignments of three or more homologous or similar sequences (DNA, protein or RNA). Table 1 shows that, typically, the alignments are represented as a two-dimensional (2-D) table, where rows represent individual sequences and columns represent residue (or base) positions.

Multiple alignments are important because they extract 'meaning' from a mass of primary sequence data. In particular, they are a concise summary of sequence relationships among homologous sequences, and they provide information on the relationship between sequences to a particular gene (protein) family, they help find weak similarities (from distantly related proteins) using only sequence data, and they generate a representative sample for all the members of the sequence family.

The 'meaning' extracted from the primary sequence data is summarized in the consensus of multiple alignments. Table 1 shows the most common representation of a consensus as a single line (or a 'pseudo sequence'), which is added at the bottom of an alignment and consists of letters that show the alignment of conserved residues (or bases) within each column. The consensus can also be represented in other ways, including Hidden Markov Models, Profiles, PRINTS, BLOCKS, regular expressions and rules. For an excellent introduction to multiple alignment and bioinformatics in general, see Attwood and Parry-Smith[3].

Table 1. Multiple alignment as a 2-D table

	Residue (or base) positions				
	1	**2**	**3**	**4**	**5**
Sequence 1	D	R	I	V	G
Sequence 2	D	R	V	V	G
Sequence 3	D	G	L	V	A
Sequence 4	D	R	L	V	G
Sequence 5	D	G	A	V	A
Consensus	D	r	–	V	g

A capital letter is a fully conserved residue, a lower case letter is a partially conserved residue.

The essential concepts of multiple alignments are that:

- There are families of sequences that share some common feature(s) – in structure or function or both – and this is reflected by the presence of conserved regions in the multiple alignment of those sequences.
- More homologous sequences in an alignment improves the chance that the sequence variation observed between them represents the variation that exists in the whole family of related proteins.
- The alignment of a family of protein sequences provides more information than the alignment of any pair of those sequences. That is, when three or more sequences are aligned, there is information in the combined sequences that is not present in any one sequence or any pair of sequences.

If information is required on the level of the relationship between sequences, or between sequences and an established sequence family, then multiple alignments must be comprised of homologous sequences. Multiple alignments can be made from similar (non-homologous) sequences to identify regions of local similarity, but then cannot be used to infer relationships between the sequences.

The many uses of multiple alignments include:

- Protein 3-D modelling
- Secondary structure prediction
- Molecular evolution – phylogeny
- Source of secondary databases (e.g. Blocks and Prints)
- Database searching
- Design of primers for PCR, drugs and vaccines
- Consensus calling in DNA sequencing
- Models for evaluation of crucial residues in enzyme action and ligand binding.

Multiple alignment methods

Figure 1 shows the general procedure for gathering the homologous sequences required for a multiple alignment. Figure 2 shows the general procedure after a multiple alignment has been generated and Figure 3 shows the two general

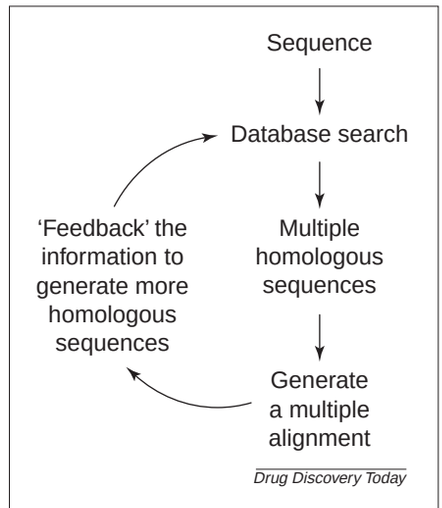


Figure 1. *Procedure for generating a multiple alignment.*

 1359-6446/00/$ – see front matter PII: S1359-6446(00)01504-X