

# Advances in the Chemistry and Pharmacology of Ecteinascidins, A Promising New Class of Anticancer Agents

I. Manzanares<sup>1</sup>, C. Cuevas<sup>1</sup>, R. García-Nieto<sup>2</sup>, E. Marco<sup>2</sup> and F. Gago<sup>2\*</sup>

<sup>1</sup>Pharma Mar S.A., Tres Cantos, E-28760 Madrid, Spain; <sup>2</sup>Departamento de Farmacología, Universidad de Alcalá, E-28871 Madrid, Spain

**Abstract:** Ecteinascidins are marine natural products consisting of two or three linked tetrahydroisoquinoline subunits and an active carbinolamine functional group. Their potent antiproliferative activity against a variety of tumor cells has made them attractive candidates for development as anticancer agents. The lead compound, ecteinascidin 743 (ET 743), is currently in phase II clinical trials but the low amounts present in its natural source, the tunicate *Ecteinascidia turbinata*, made it necessary to develop efficient synthetic procedures. Recent improvements on the original synthesis are reviewed as well as new strategies starting from readily available cyanosafrafrin B. ET 743 is known to bind to the minor groove of DNA giving rise to a covalent adduct with the exocyclic amino group at position 2 of a guanine in a fashion similar to saframycin antibiotics. Some of the resulting complexes have been studied by a variety of biochemical and spectroscopic methods and also by computer simulations. The rules for sequence specificity have been well established (preferred targets are RGC and YGG, where R and Y stand for purine and pyrimidine, respectively), and it has been shown that binding of ET 743 to DNA is accompanied by minor groove widening and DNA bending towards the major groove. Although the precise target for antitumor action remains to be unambiguously defined, a role in affecting the transcriptional regulation of some inducible genes is rapidly emerging.

## INTRODUCTION

### Natural History

*Ecteinascidia turbinata* is a tunicate species from the Caribbean and the Mediterranean that belongs to the class Ascidiacea within the subphylum Tunicata (also called Urochordata). Ascidiaceans, or sea squirts, are small bottom-dwelling soft-bodied marine animals that form colonies comprising many individuals, called zooids. The name tunicate derives from their characteristic protective covering, or tunic, which functions to a certain extent as an external skeleton and consists of some cells, blood vessels, and a secretion of a variety of proteins and carbohydrates, including cellulose, an unusual finding in animals. Within the tunic is the muscular body wall, which controls the opening of the siphons used for feeding. Colonies are formed by asexual reproduction through budding, that is, outgrowths on the parent break off as new individuals. Sexual reproduction, on the other hand, leads to a fertilized egg that develops into a free-swimming tadpole larva. The larval stage is brief and is used to find an appropriate place for the adult to live. Larvae have notochords and nerve cords, as well as muscular tails twice as long as their bodies. Following attachment to a surface, the body resorbs the tail, using it as a food supply during metamorphosis into its sessile adult form. Degeneration of the dorsal nerve cord and sensory organs leave in the adult a single ganglion between

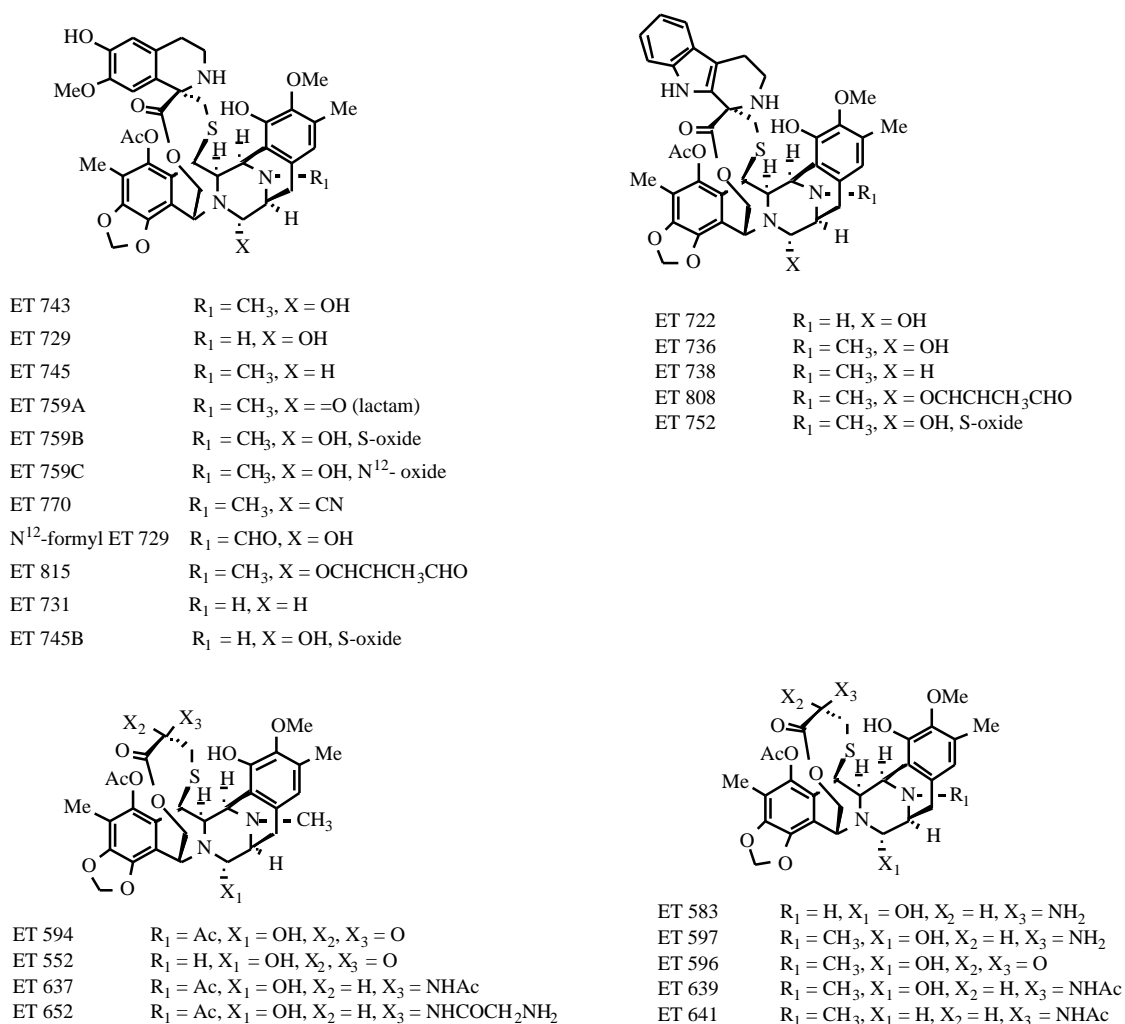
the oral and atrial openings from which nerves grow to the various organs of the body.

### BIOLOGICAL ACTIVITY AND CHARACTERIZATION OF THE ACTIVE COMPOUNDS

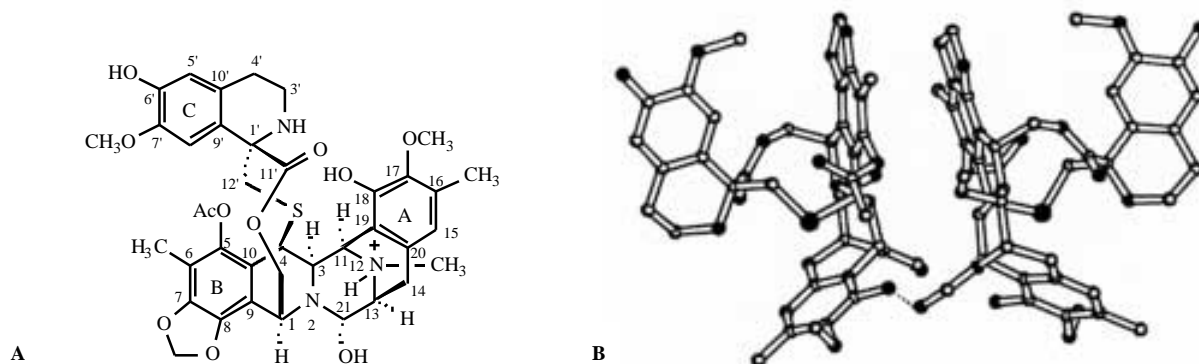
Crude aqueous ethanol extracts of *Ecteinascidia turbinata* were shown to have powerful immunomodulating and antiproliferative properties as early as 1969 [1]. Many unsuccessful attempts at isolation of the active compounds were followed by concerted efforts that led to the characterization of six alkaloids called ecteinascidins 729, 743, 745, 759A, 759B, and 770, of which ET 743 was the most abundant (0.0001% yield) [2]. The numbers after the abbreviation ET (Fig. 1) refer to the masses (M) deduced from the highest mass ions (M + H) observed in positive-ion Fast Atom Bombardment Mass Spectrometry (FABMS) spectra. Tandem FABMS/MS was particularly useful in defining structural units in the compounds, as discussed in the recent review by Kenneth L. Rinehart [3].

The novel and unique chemical structure of ecteinascidins (Fig. 2) is formed by a monobridged pentacyclic skeleton composed of two fused tetrahydroisoquinoline rings (subunits A and B) linked to a 10-member lactone bridge through a benzylic sulfide linkage. Most ecteinascidins have an additional tetrahydroisoquinoline or tetrahydro-carboline ring (subunit C) attached to the rest of the structure through a spiro ring. The existence of these two families reveals different biosynthetic pathways very likely involving condensation of carbonyl groups with either dopamine or tryptamine equivalents [3,4].

\*Address correspondence to this author at the Departamento de Farmacología, Universidad de Alcalá, E-28871 Madrid, Spain; Tel: +34 - 918 854 514; Fax: +34 - 918 854 591; E-mail: federico.gago@uah.es



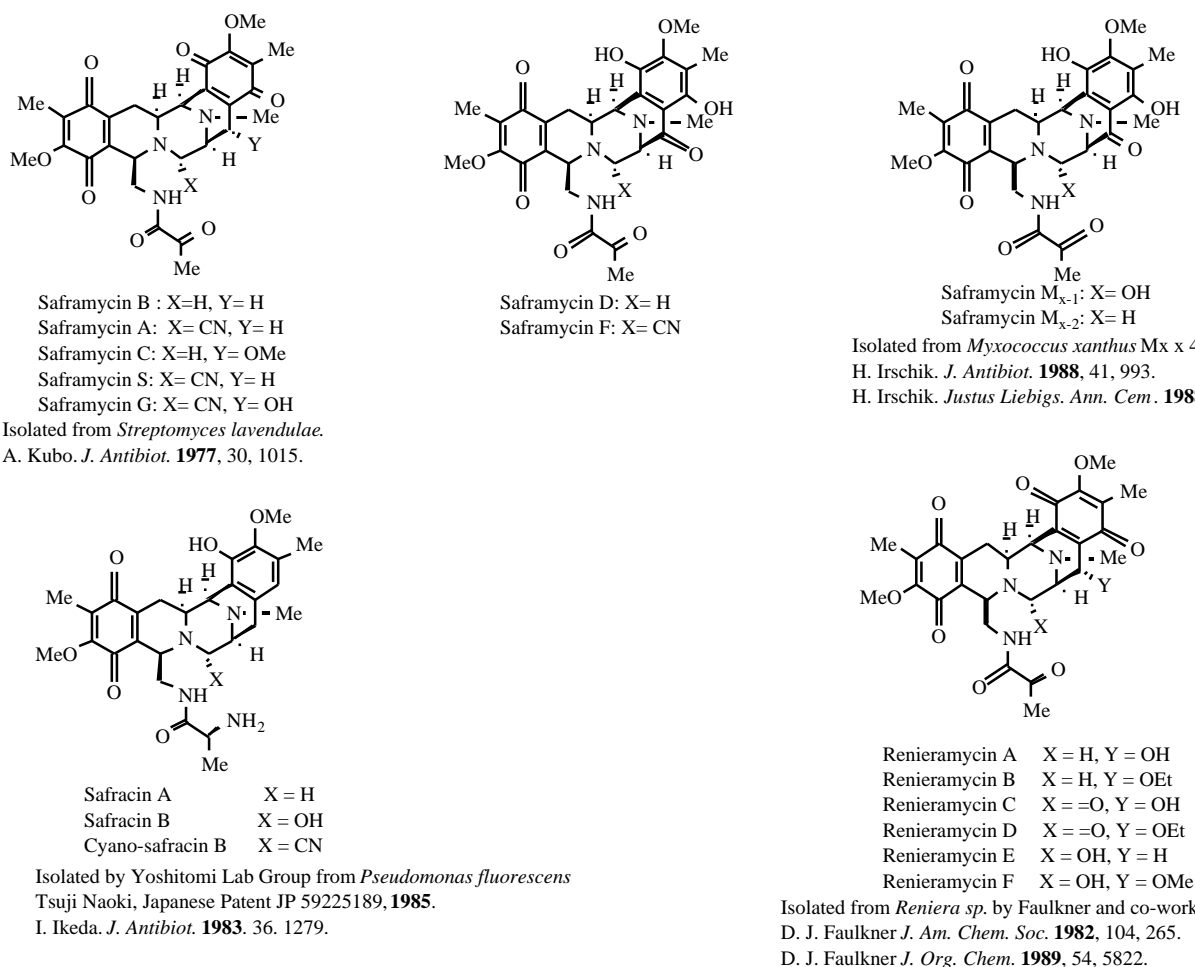
**Fig. (1).** Chemical structures of ecteinascidins.



**Fig. (2).** (A) Chemical formulae and numbering of ET 743. (B) Association of two ecteinascidin molecules as found in the crystal structure of N<sup>12</sup>-formyl C<sup>21</sup>-methoxy ecteinascidin 729 [11]. The dotted line represents a hydrogen bond between the phenol oxygen at C18 in subunit A of one molecule and the formyl oxygen from a neighbouring molecule.

A clear structural similarity (Fig. 3) is apparent with microbially derived safracins [5] and saframycins [6], both less potent antitumour agents than ecteinascidins, and also with sponge-derived renieramycins and xestomycin, for which no antitumour activity has been reported [3]. The

antitumour action was early thought to be largely due to the presence of an active  $\alpha$ -carbinolamine (N-C-OH) group [7], which is also present in naphthyridinomycins, pyrrolo[1,4]benzodiazepine antibiotics such as anthramycin, sibiromycin and tomaymycin, and the quinocarcins.



**Fig. (3).** Chemical structures of safracins, saframycins, and renieramycins.

Reduction of the quinone skeleton of saframycin A to hydroquinone results in the concomitant elimination of the nitrile group and enhanced covalent binding to DNA [8], in a reaction reminiscent of that between anthramycin and DNA [9]. An  $\alpha$ -carbinolamine or an iminium were the species suggested to be involved in the interaction, and the amino group of guanine in the minor groove was later identified as the site of covalent attachment [10]. By analogy to these related antibiotics, the potent biological activity of ecteinascidins was then rapidly associated with their ability to form a covalent adduct with DNA using the reactive carbinolamine group [11].

The chemical structures of the ecteinascidins were assigned by correlation NMR techniques and FABMS spectra whereas the unequivocal assignment of the relative stereochemistries at C-4 and C-22 (=C1') was achieved only when the X-ray crystal structures of the natural  $N^{12}$ -oxide of ET 743 and the 21-*O*-methyl- $N^{12}$ -formyl derivative of ET 729 were solved [11,12].

#### X-RAY CRYSTALLOGRAPHY AND PROTONATION STATE OF ECTEINASCIDINS

The crystal structures of  $N^{12}$ -formyl C<sup>21</sup>-methoxy ET 729 formic acid methanol solvate dihydrate and ET 743  $N^{12}$ -

oxide acetonitrile solvate octahydrate (deposited in the Cambridge Structural Data Base under accession codes WAMBAN and WAMBER, respectively) revealed two independent ecteinascidin molecules per asymmetric unit plus several solvent molecules. The molecular shape of both ecteinascidins is highly compact, with much of the surface occupied by hydrophobic methyl, methylene and aryl-CH hydrogens, and only the hydroxyl groups at C18 of subunit A and C27 (=C6') of subunit C, which point away from the molecule, available for hydrogen bonding.

The ecteinascidin molecule was first described as a platform formed by the B subunit ring system on which both subunits A and C stand using their cross-bridging skeletons. In the crystals, the two molecules are associated into dimers of ellipsoidal shape through the stacking of their ring B systems and formation of an intermolecular hydrogen bond involving the hydroxyl group on C18 of one molecule and either the formyl or the *N*-oxide oxygen on N12 of the other molecule (Fig. 2B). This hydrogen bond appears to provide a swivel point for the association. The presence of formate ion in the WAMBAN crystal structure was suggestive of protonation of one of the nitrogen atoms. Since N2 is almost totally shielded from the solvent, N12 was proposed as the most likely candidate, in analogy to saframycin, which becomes selectively protonated at the N12 position at pHs

below 5.5 [10]. Later, proof was obtained by NMR spectroscopy that N12 is indeed protonated in the covalent DNA adduct as well [13,14].

## ANTIPROLIFERATIVE PROPERTIES AND STRUCTURE-ACTIVITY RELATIONSHIPS

Initial structure-activity relationships among natural members of the ecteinascidin class were based on results from *in vitro* cytotoxic screening assays against murine leukemia cells L1210 [15]. The importance of the reactive carbinolamine for optimal activity was highlighted by the 175-fold decrease in activity in going from ET 743 to its C-21-deoxy counterpart, ET 745, and the 17-fold reduction in potency upon oxidation of the carbinolamine to the lactam (ET 759A). Likewise, replacement of the C-21 hydroxyl group by a malonaldehyde (ET 815) or a cyano (ET 770) group reduced the cytotoxic potency relative to ET 743, most likely due to slower formation of the reactive iminium intermediate. Other structural modifications were also shown to influence cytotoxicity. Thus, deacetylation of ET 743 to give ET 701 was seen to reduce the activity 40 times whereas the *N*-12 demethylated analog, ET 729, was 10 times more active than ET 743. Acetylation or methylation of the free phenol group in subunit C, on the other hand, did not significantly alter the biological properties of ET 743 but the activity was reduced 40 times when the phenol group at position C-18 (subunit A) was methylated. In general, the presence of the aromatic C subunit was found to be important for potency, as exemplified by comparing ET 743 and ET 729 with ET 583, ET 594 and ET 597, but little difference in activity was found between ecteinascidins containing a tetrahydroisoquinoline ring system as subunit C (*e.g.* ET 729 or ET 743, and the corresponding analogs containing a tetrahydro-*β*-carboline ring (*e.g.* ET 722 or ET 736).

*In vitro* cytotoxicity studies with L1210 mouse leukemia cells were then extended to other cell lines, and subnanomolar potencies were also established against P388 mouse leukemia, A549 lung cancer, HT29 colon cancer, MEL-28 melanoma cells and human tumours explanted from patients [12]. *In vivo* activity was then evaluated in mouse tumour models and a variety of human tumours xenografted in nude mice. Complete regressions were observed in MEXF989 melanoma, MX-1 breast carcinoma, LXFL529 non-small cell lung carcinoma and HOC22 ovarian carcinoma xenografts, and partial regressions in renal MRIH121 and PC2 prostate carcinoma xenografts [3]. Interestingly, changes in the C subunit were seen to modulate the biological activity of both ET 736 and its N12 demethyl analogue, ET 722, relative to ET 743 suggesting that this part of the molecule plays an important role in cytotoxicity. For example, ET 722 and ET 736 displayed a higher level of activity *in vivo* against P388 leukemia in mice whereas ET 729 and ET 743 showed higher activity against B16 melanoma, Lewis lung carcinoma, M5076 ovarian sarcoma, and MX1 human mammary carcinoma xenografts [12]. ET 743 was then selected for clinical development because of its greater relative abundance in the tunicate, and entered clinical trials in early 1996 both in Europe and in the United States. ET 743 is now in phase II clinical trials and toxicity

so far has been shown to follow a transient-reversible pattern and to be dose-related, predictable and mostly limited to bone marrow and liver [16].

The observation that ET 743 analogues lacking subunit C and possessing various substituents on the nitrogen atom of the  $\alpha$ -amino lactone still retained good antitumour activity led to the design and synthesis of a number of ecteinascidin-like compounds [17]. Replacement of the 10-membered lactone bearing the C subunit with benzamide, *cis*-cyclohexane-1,2-dicarboximide, *cis*-cyclohexene-1,2-dicarboximide or phthalimide moieties, together with protection of the phenolic hydroxyl group on the B subunit by acetyl, propionyl, methoxyacetyl, methansulfonyl, methyl, or ethyl groups, yielded a number of derivatives. Their primary antiproliferative activities were tested *in vitro* using different human solid cancer cell lines. After 3-day continuous exposure to the drugs a metabolic assay was used in which the cellular reduction of a tetrazolium salt afforded a colorimetrically detectable formazan in proportion to viable cell number. These studies confirmed and expanded the initial structure-activity relationships and showed that an acetoxy group in the B subunit afforded optimum activity whereas protection of the other phenolic hydroxyl group in the A subunit resulted in diminished antitumour activity. The most potent compound turned out to be phthalascidin (Pt 650), which displayed antiproliferative activities comparable to those of ET 743, and in both cases independent of cellular growth rate and *p53* tumor suppressor gene status. The dose-response curves were also similar, including the intriguing observation that cell viability after 24 h of drug exposure appeared to be 10-20% greater at 10-100 nM than at 1-2 nM, a finding suggestive of a relatively narrow concentration range for optimum bioactivity.

## SYNTHETIC STUDIES

*Ecteinascidia turbinata* has been successfully grown by Pharma Mar in aquaculture facilities in Spain (near Formentera island, in the Mediterranean sea) in what represents a more practical and environmentally sound practice than harvesting the creature from the wild. Nevertheless, in recent years several synthetic schemes have been developed that make it possible to produce ET 743 in the kilogram quantities required for further clinical studies worldwide.

### Synthetic Routes to the Saframycins

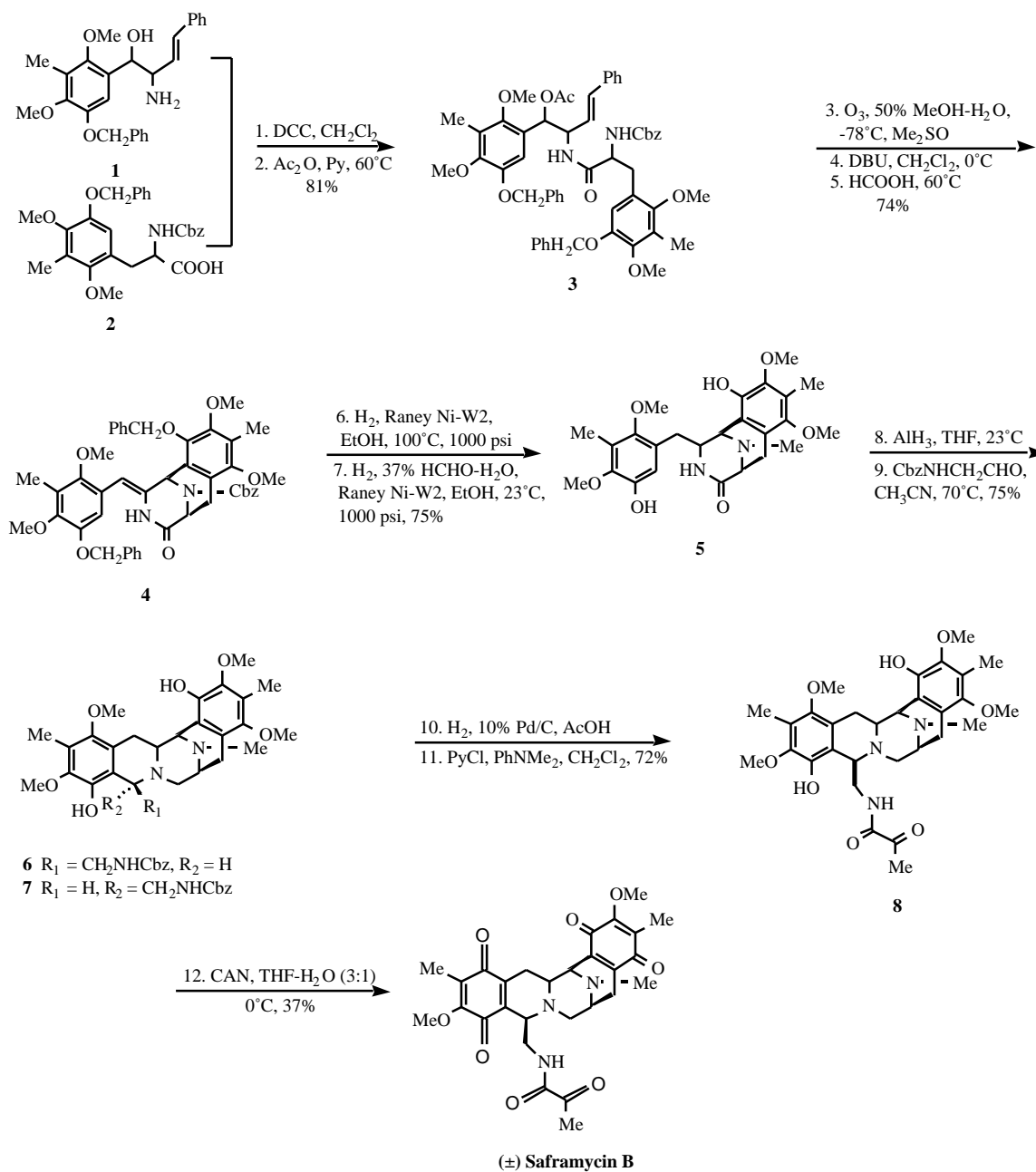
The precedents to the first total synthesis of ET 743 are to be found in synthetic work leading to related molecules such as saframycins, safracins and renieramycins, the structures of which are shown in (Fig. 3). These antimicrobial compounds, isolated from bacterial sources or marine sponges, have the same pentacyclic skeleton as the ecteinascidins but one or two of the aromatic rings contain a different oxidation pattern, resulting in dimeric mono- or bis-quinone structures.

Synthetic approaches to members of the saframycins family started in 1982 with the total synthesis of ( $\pm$ )-saframycin B, achieved by T. Fukuyama *et al* [18] (Scheme

1). In this approach the amide intermediate **3** was obtained by condensation of amine **1** and acid **2**, and subsequent acetylation. The bicyclic [3.3.1] system **4** was produced by crucial double cyclization in a three-step sequence. Ozonolysis of **3** followed by treatment with dimethyl sulfide produced a diastomeric mixture of aldehydes. Acetate elimination with DBU gave a mixture of the *cis-trans* enamines. On heating in formic acid, this mixture was exclusively converted to the desired bicyclic **4** in 74% overall yield from **3**. This highly selective cyclization can be explained in terms of rapid isomerization through protonation-deprotonation of the intermediate enamine

formed. Subsequent catalytic hydrogenation of **4** and reductive alkylation yielded the *N*-methylamine **5**. Reduction of the lactam **5** to the amine followed by phenolic cyclization gave compound **6** and the epimer **7**, in a 6:1 ratio, respectively, and 75% yield. Deprotection of the carbobenzyoxy group and subsequent acylation with pyruvoyl chloride provided the pyruvamide **8**. Oxidation of the phenol **8** using CAN gave ( $\pm$ )-saframycin B.

Later on, A. Kubo *et al* [19] reported the total synthesis of ( $\pm$ )-saframycin B using arylidenepiperazinedione **9** as a key intermediate. The synthetic plan is outlined in Scheme



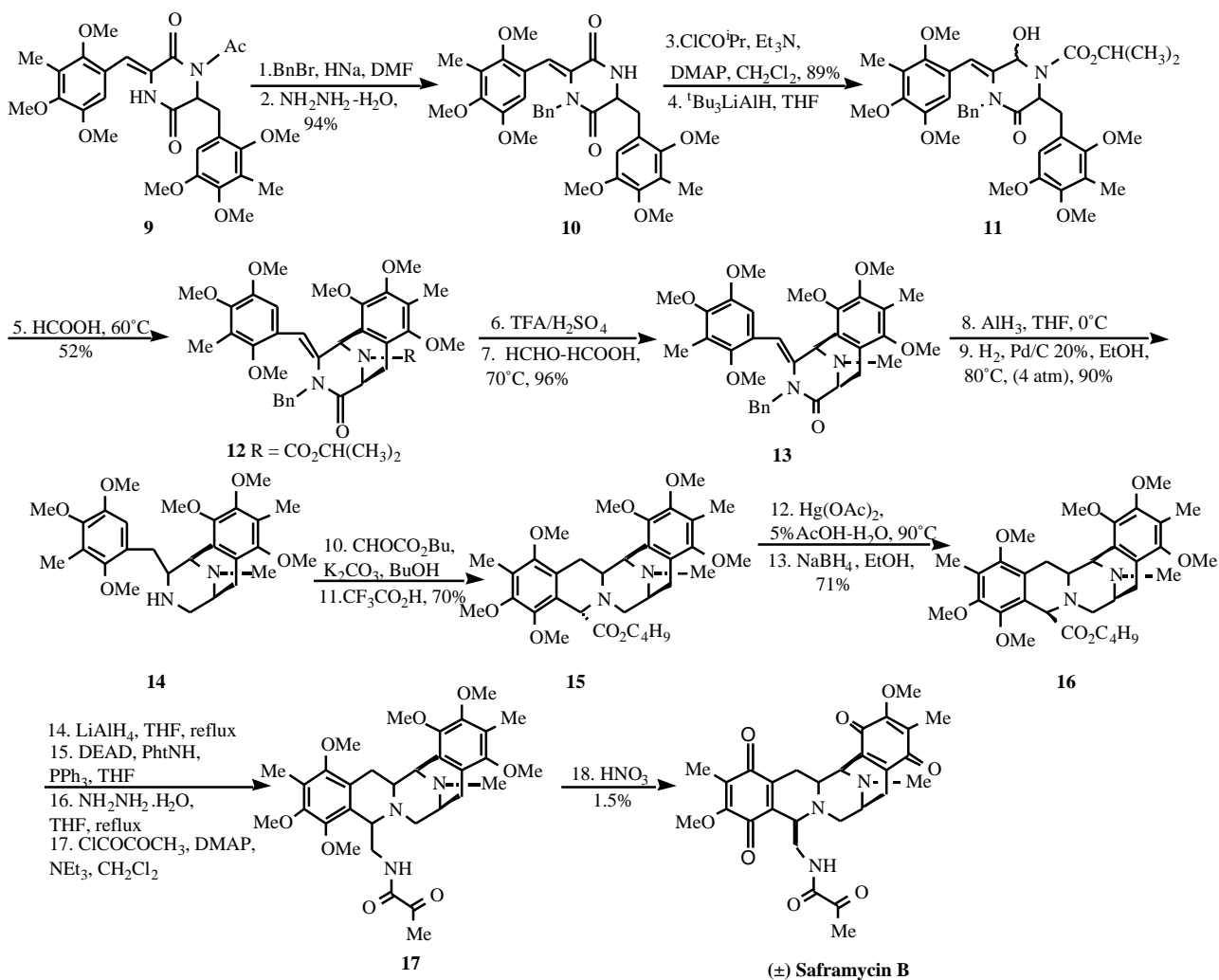
Scheme 1.

(2). Key transformations involve the regioselective five-step synthesis of derivative **12**, and subsequent conversion to the secondary amine **14**. Benzoylation of **9** with benzyl bromide followed by deacetylation with hydrazine hydrate afforded **10**. Subsequent acylation and reduction with tri-*tert*-butoxyaluminum hydride gave the allylic alcohol **11**. The cyclization of **11** by treatment with formic acid at 60 °C yielded the 1,5-imino-3-benzazocine **12**, which is the first key intermediate in the synthetic plan. The conversion of **12** into secondary amine **14** achieved by deprotection of **12** with TFA-H<sub>2</sub>SO<sub>4</sub> and methylation with formaldehyde and formic acid at 70 °C. The tricyclic lactam **13** obtained was transformed with lithium aluminum hydride, methylation and subsequent reduction of the double bond afforded the amine **14**. A modified Pictet-Spengler cyclization gave the pentacyclic intermediate **15**. The oxidation of **15** followed by reduction gave compound **16**. Reduction of the ester **16**, subsequent reaction with diethyl azocarboxylate and phthalimide, followed by deprotection, afforded the amine which was coupled with pyruvoyl chloride to give

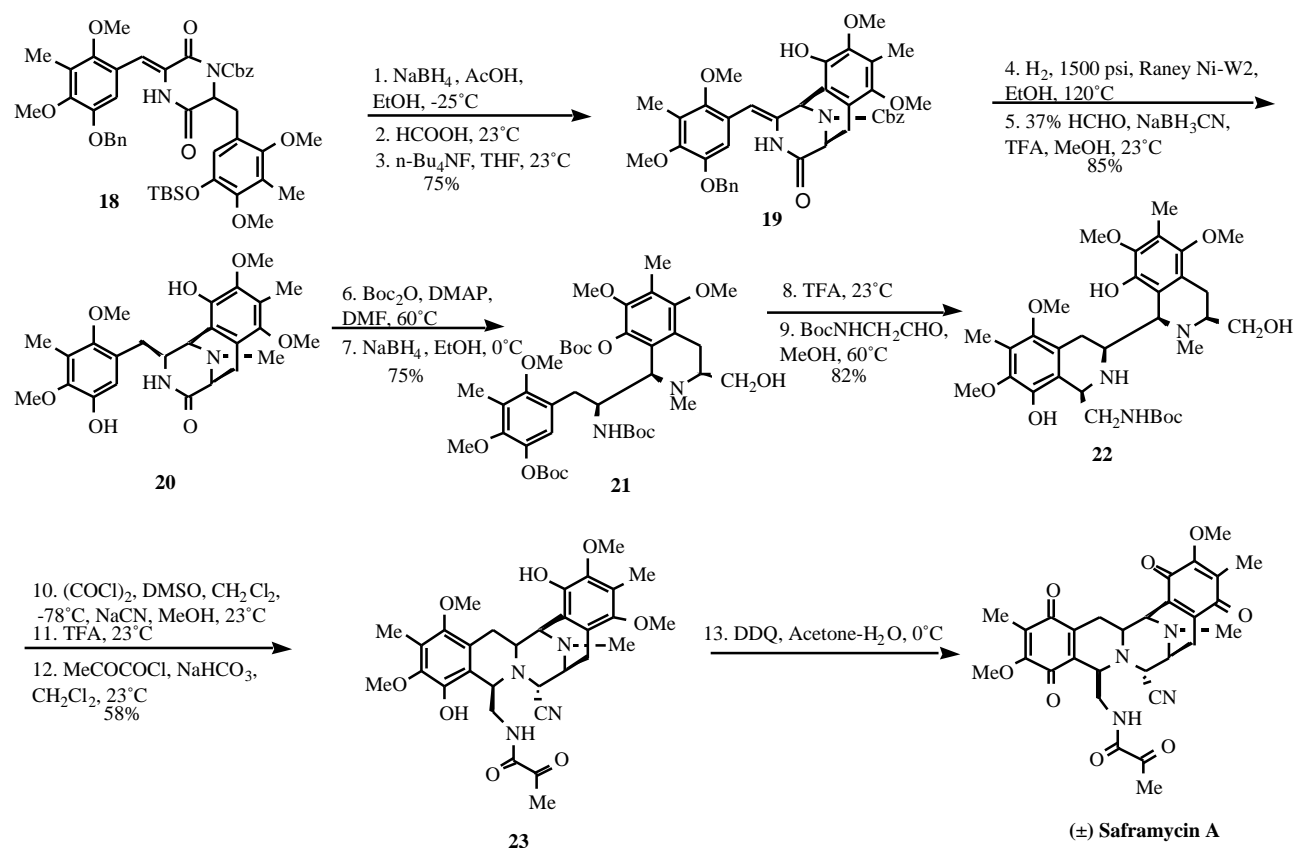
pyruvamide **17**. The oxidation with HNO<sub>3</sub> afforded (±)-saframycin B in 1.5% yield.

Related synthetic studies were performed by A. Kubo *et al* on the synthesis of saframycin A [20], in the transformation of (±)-saframycin B to (±)-saframycin C and D [21], and in the transformation of (–)-saframycin A to (–)-saframycin Mx [22], using SeO<sub>2</sub> in different reaction solvents (Scheme 2).

The total synthesis of (±)-saframycin A was reported by T. Fukuyama *et al* [23], who used the arylidenpiperazine-dione **18** as the key intermediate, as outlined in Scheme (3). The selective reduction of the activated ring carbonyl group, facile acyliminium ion-mediated cyclization, and subsequent deprotection afforded the bicyclic[3.3.1] system **19**. Catalytic hydrogenation of **19** gave the diphenol amine which, upon reductive methylation, yielded the methyl amine **20**. Cleavage of the lactam present in **20** to obtain **21** was facilitated by protection with the Boc group. Deprotection of



Scheme 2.



Scheme 3.

the Boc groups and Pictet-Spengler cyclization yielded **22**. Oxidation of **22** and subsequent treatment of the unstable iminal with NaCN furnished the amino nitrile which was transformed into the pyruvamide **23**. Oxidation of **23** with DDQ afforded (±)-saframycin A in 60% yield. The total synthesis of (±)-renieramycin A was also accomplished following a very similar pathway [24].

More recently two enantioselective synthesis of (–)-saframycin A have appeared. Myers *et al* [25] achieved an enantioselective synthesis by direct condensation of *N*-protected  $\alpha$ -amino aldehyde **24** and the *C*-protected  $\alpha$ -amino aldehyde **25** (Scheme 4). Coupling of **24** and **25** produced imine **26**. Treatment of **26** with a saturated solution of anhydrous lithium bromide and warming to  $35^\circ\text{C}$  brought about a Pictet-Spengler cyclization to give a 5:1 mixture of *cis-trans* tetrahydroisoquinolines. Flash column chromatography of this mixture afforded **27** in 65% yield. Reductive methylation and deprotection of the protecting groups provided **28** in 82% yield. Addition of *N*-Fmoc glycinal to amine **28** produced an imine intermediate that underwent Pictet-Spengler cyclization affording compound **29**. Deprotection of the *N*-Fmoc group by treatment with DBU and subsequent coupling of the amine with pyruvoyl chloride gave the pyruvamide **30** which was oxidized with iodobenzene furnishing (–)-saframycin A. Later, Myers *et al* [26] assembled the entire skeleton of saframycin A in one low-yield transformation from an *N*-linked oligomer of three  $\alpha$ -amino aldehyde components.

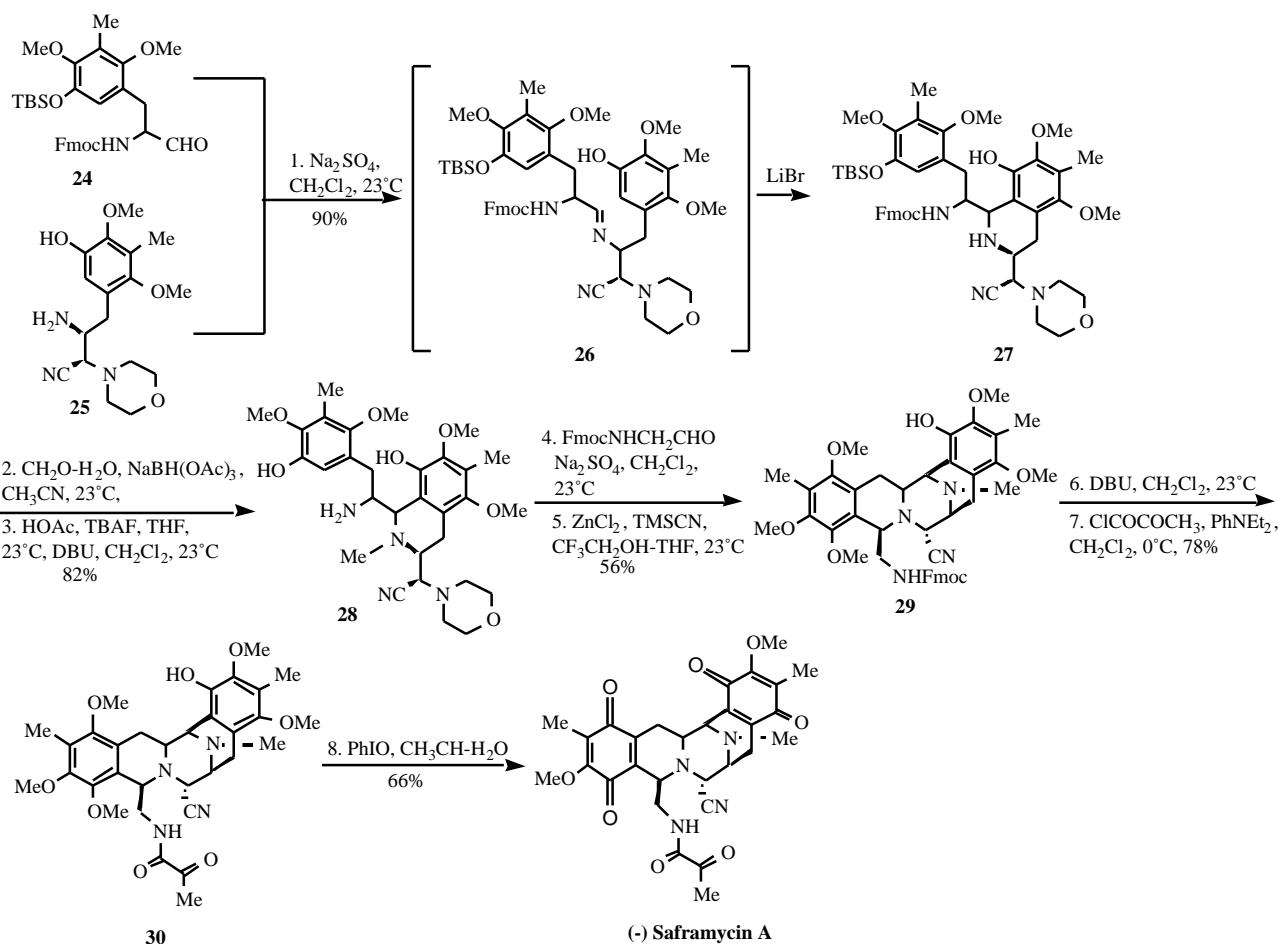
Another enantioselective synthesis of (–)-saframycin A was described by E. J. Corey *et al* [27] using an intermediate, **37**, previously employed in the total synthesis of ET 743 (Scheme 5). In this pathway the phenolic function of **37** was protected and the alcohol group was deprotected and transformed into the pyruvamide **31** via the corresponding azide. Deprotection of the allyl group and additional oxidation afforded the 1,4-benzoquinone **32**. Selective *O*-methylation and subsequent catalytic oxidation provided (–)-saframycin A.

Finally, Danishefsky *et al* carried out enantioselective studies on the saframycin series constructing the major chiral subunits through catalytic asymmetric epoxidation and asymmetric dihydroxylation [28].

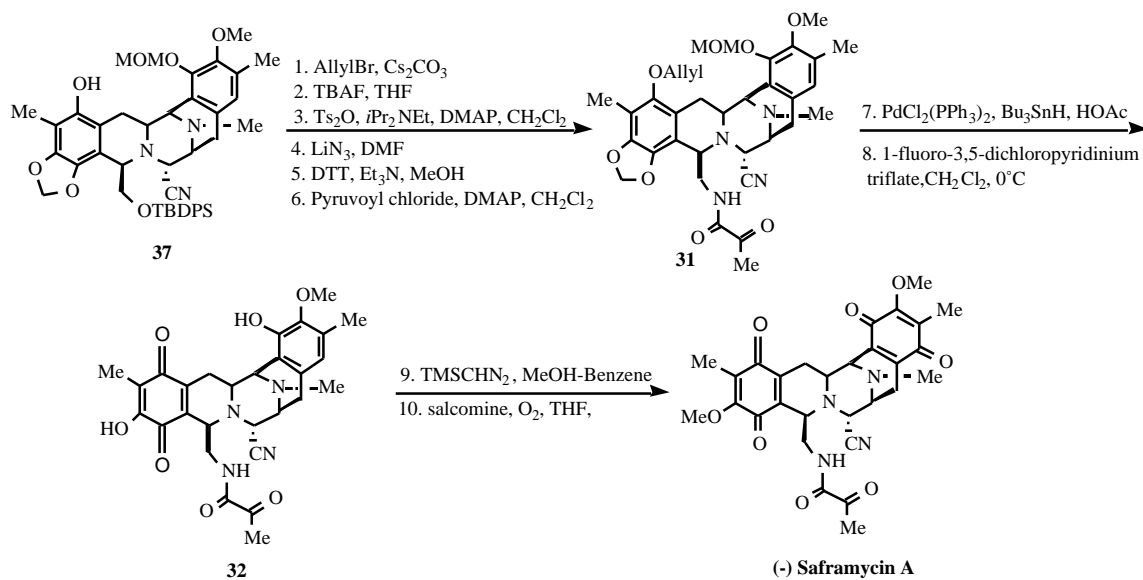
### Synthetic Routes to Ecteinascidins

To date two synthetic routes to ET 743 have been reported. E. J. Corey *et al* achieved the first enantioselective total synthesis of ET 743 in 1996 [29], and four years later, I. Manzanares *et al* reported the complete hemisynthesis of ET 743 starting from readily available cyanosafraicin B [30].

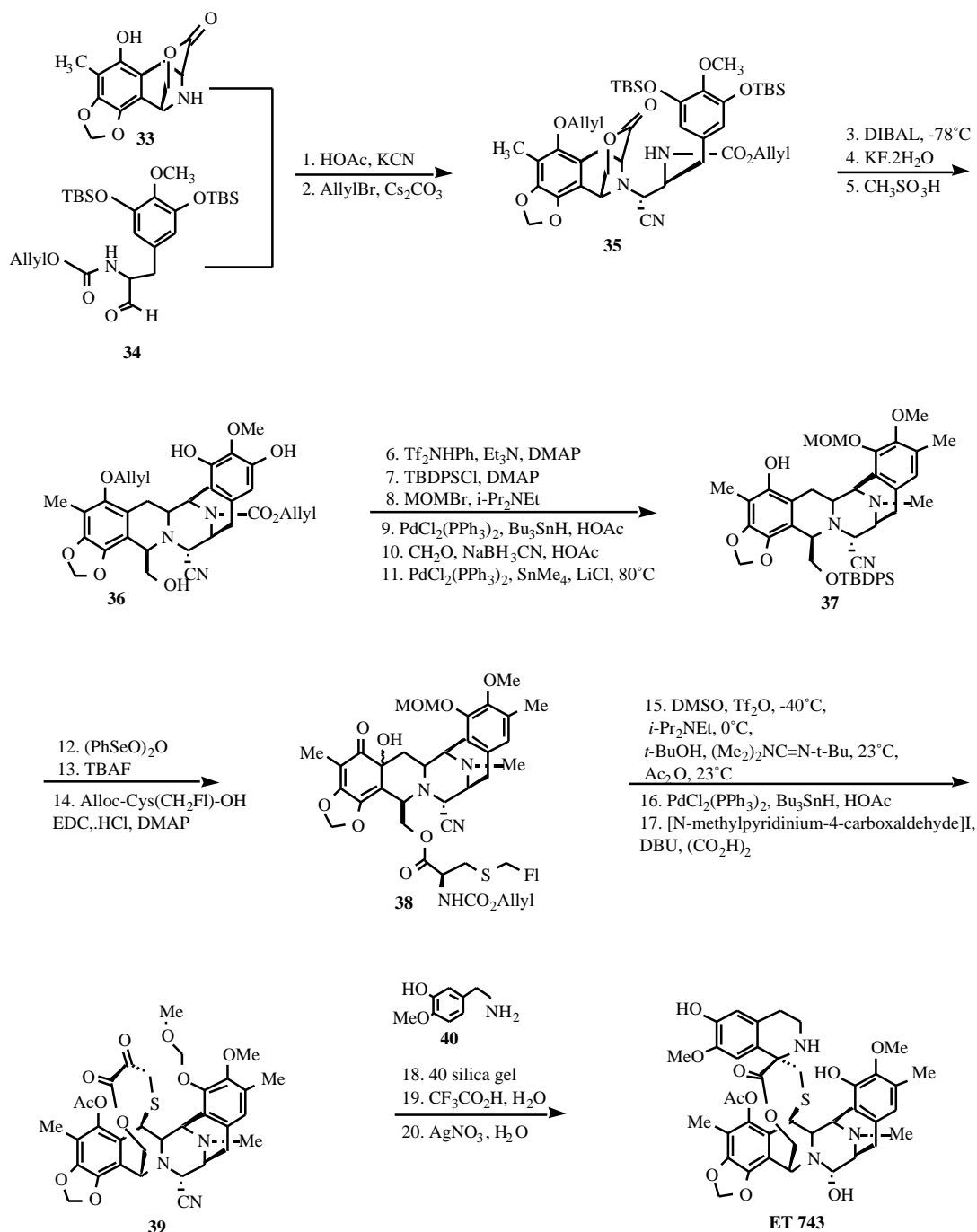
The synthesis by Corey *et al* is described in Scheme (6). The enantioselective syntheses of bridge lactone intermediate **33** and *N*-protected  $\alpha$ -amino aldehyde **34** using catalytic asymmetric hydrogenation provided two important



Scheme 4.



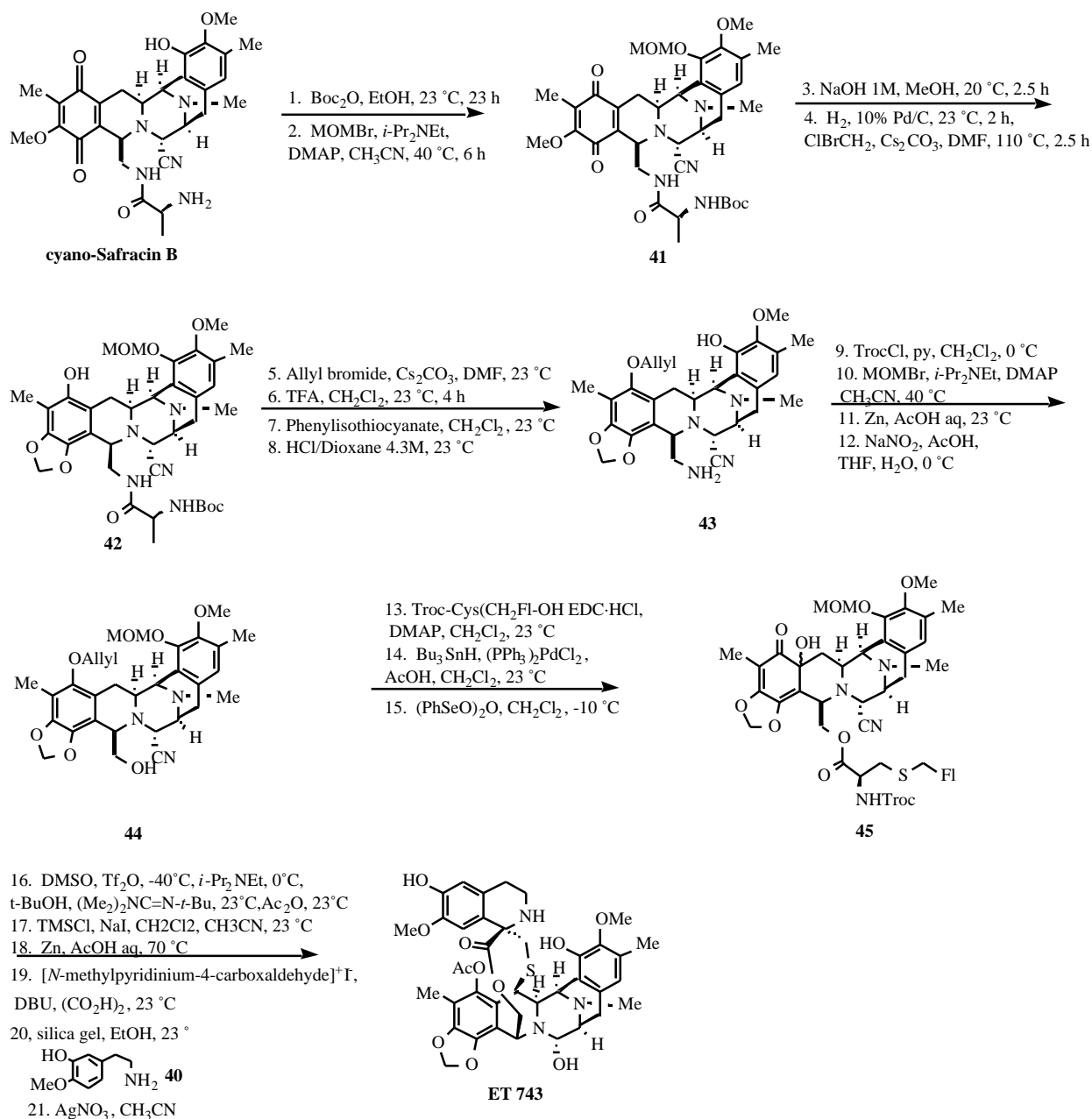
Scheme 5.



Scheme 6.

intermediates to build the piperazine-bridged bis (tetrahydroisoquinoline) framework. Coupling of building blocks **33** and **34**, reduction of intermediate lactone **35** to the corresponding lactol, and internal Mannich bisannulation gave the monobridged pentacyclic intermediate **36**. After protection, *N*-methylation and replacement of CF<sub>3</sub>SO<sub>3</sub> by CH<sub>3</sub>, intermediate **37** was obtained, which has the required

substitution pattern in the pentacyclic framework. From intermediate **37** the following key transformations completed the synthesis of ET 743: (1) oxidation of phenol **37** permitted the selective hydroxylation of the angular position of the aromatic ring; (2) after esterification of the primary hydroxyl function with a protected cysteine derivative, compound **38** was transformed in a one-pot reaction into the 10-membered



Scheme 7.

lactone bridge by creation of exo quinone methide followed by nucleophilic addition of de-protected cysteine and further acetylation of the resulting phenoxide ion; (3) deprotection of the Alloc group and subsequent transamination of the -amino lactone gave the -keto lactone **39**; (4) Pictet-Spengler cyclization with phenethylamine **40** generated the spiro tetrahydroisoquinoline stereospecifically.

Later, E. J. Corey *et al* published a modified procedure that is considered more effective than the previous one [17]. The reported improvement involved formation of a **35**-like intermediate in the initial coupling but with the CN group replaced by a carbonyl moiety.

The synthesis by Manzanares *et al* [30] is described in Scheme (7). Starting from cyano-safracin B it was possible to achieve the synthesis of ET 743 in a very short and straightforward way. Safracin B is available through fermentation of the bacteria *Pseudomonas fluorescens*, and optimization of the fermentation process has allowed the preparation of the cyano derivative on a kilogram scale. Selective hydrolyzation of protected methoxy-*p*-quinone **41**, reduction of the quinone and reaction of the reduced compound with bromochloromethane afforded the methylenedioxy intermediate **42**. Protection of **42** with allyl bromide, deprotection of the MOM and Boc groups, and subsequent cleavage of the amide bond by Edman's

degradation gave the amine **43**. Protection of the phenolic function in **43** and conversion of the primary amino group into an alcohol afforded intermediate **44**. Esterification of compound **44**, allyl deprotection and angular oxidation provided intermediate **45**. The synthesis was completed in six additional steps. Salient features of this end-game operation was the early removal of the MOM group with TMSI and the use of Troc protection for the amino function of the (*S*)-cysteine derivative.

This novel methodology has resulted in a substantial improvement in the overall yield of the synthesis of ET 743 with respect to the earlier Corey synthesis.

### Synthetic Routes to Phthalascidin

The ET 743 analog phthalascidin (Pt 650) was originally described by E. J. Corey *et al* [17] as a compound with comparable activity to ET 743. Pt 650 was synthesized (Scheme 8) using the common pentacyclic **36** (Scheme 6) obtained in the synthetic route to ET 743. The pentacyclic triol **36** was converted to the phenolic monotriflate **46**. Protection of the alcohol and phenol groups gave compound **47**. Cleavage of the *N*-Alloc and *O*-allyl group in **47** afforded the secondary amine **48**, which was *N*-methylated and *C*-

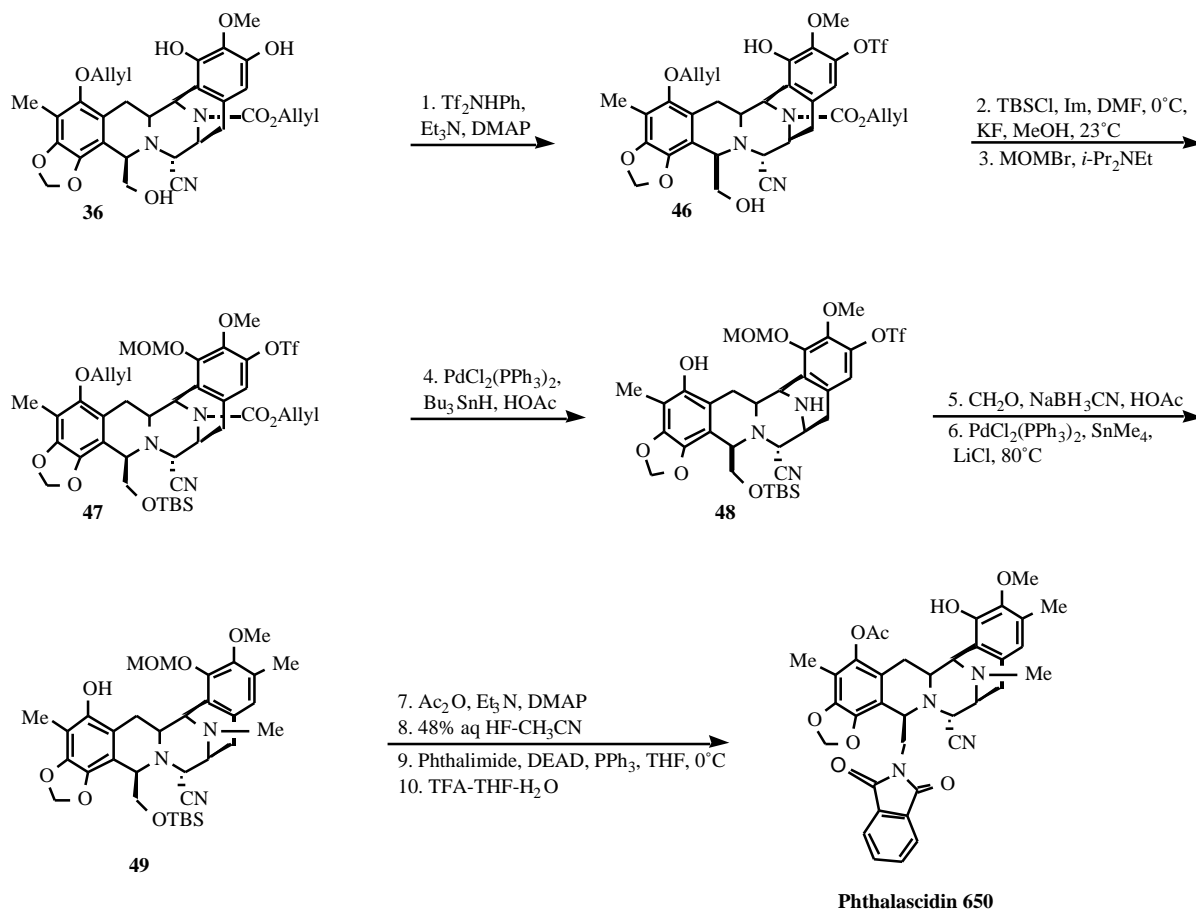
methylated to **49**. Acetylation of phenol **49**, deprotection of the silyl group, and Mitsunobu displacement of the primary alcohol produced the phthalimide, which upon acid catalyzed cleavage of the methoxymethyl ether provided pure Pt 650.

More recently, Manzanares *et al* [30] reported the synthesis of this compound (outlined in Scheme 9) starting from intermediate **42**, previously utilized in the synthesis of ET 743 from cyanosafracin B. Acetylation of **42**, deprotection of the *N*-Boc and MOM groups followed by Edman's degradation of the alanine side chain afforded **50**, which was allowed to react with phthalic anhydride and carbonyldiimidazole to give phthalascidin. This synthetic approach can be easily extended to the preparation of a wide range of analogs.

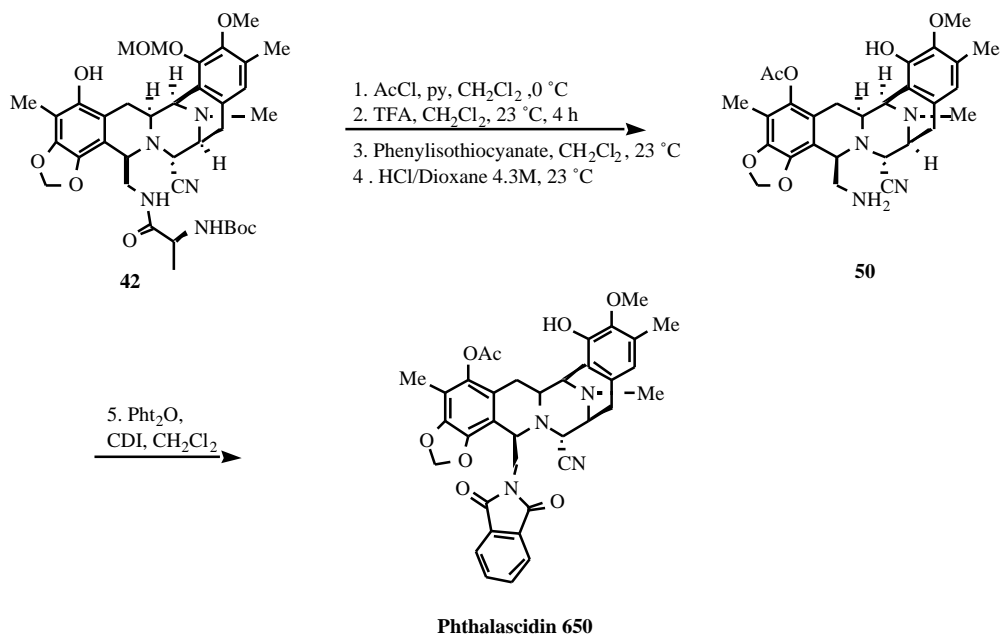
### BIOCHEMICAL AND STRUCTURAL STUDIES ON ECTEINASCIDINS

#### Structural Basis for Reactivity and Sequence Specific DNA Binding

Following determination of the crystal structure of the ecteinascidin derivatives, a molecular modelling study addressed the question of sequence specificity and the mode



Scheme 8.

**Scheme 9.**

of interaction of ET 729 and ET 743 with DNA [11,12] Given the distinctive pattern of hydrogen bond acceptors and donors, and the fact that guanine is alkylated on the exocyclic amino group, critical spatial positioning between the active carbinolamine functional group (N2-C21-OH) of the drug and the N2 of guanine appeared crucial for the cross-linking reaction in the DNA minor groove. As a result of this study, both an orientation for bound ET and a role for hydrogen bonding in the sequence recognition of DNA by these drugs were suggested: subunits B and C would stack on the sugar-phosphate backbone each on one side of the DNA minor groove, following the right-handed curvature of the double helix, and a number of intermolecular hydrogen bonds involving N2 of guanines would stabilize the complex. Accordingly, GGG was proposed to be the best sequence for complex formation. It is remarkable that, despite the bulkiness of the ET molecule, the minor groove was reported to be minimally disturbed but this was probably a consequence of using only a steepest descent energy minimization algorithm in model refinement.

Direct evidence that ET 743 alkylates duplex DNA at guanines was provided by a series of gel electrophoresis experiments [31] which showed (i) no detectable effect on single-stranded DNA, (ii) retarded migration of both supercoiled and nicked simian virus 40 (SV40) DNA following incubation with increasing concentrations of ET 743, (iii) strong retardation with a homopolymeric duplex decanucleotide containing dG:dC but no discernible effects with decamers containing either dA:dT or dI:dC, (iv) progressive covalent adduct formation following noncovalent binding of the drug, and (v) reversion of ET 743-induced DNA mobility effects upon DNA denaturation. DNA footprinting experiments using DNase I and 1,10-

phenantroline-copper plus hydrogen peroxide (both of them reactives that attack the DNA from the minor groove), and also exonuclease III (which digests DNA from its 3'-ends and is stopped by DNA adducts) revealed a protection by ET 743 on the minor groove of three to five bases and a sequence preference different from that of anthramycin. As expected, no effects were detectable when guanine was replaced by inosine in the *Escherichia coli* tyrosine tRNA promoter (*tyrT* DNA) used in the experiments. To gain further insight into the sequence selectivity, a series of duplex 14mer oligonucleotides containing a central guanine flanked on each side by different combinations of bases (5'-XGY) were studied in a band shift assay. Outside the central triplet all guanines were replaced with inosine to eliminate possible ambiguities. ET 743 was shown to retard the DNA in a concentration-dependent manner and to produce a more pronounced shift when compared to anthramycin. When all central triplets were compared to 5'-GGG (run as a standard for calculation), alkylation by 10 and 100  $\mu$ M ET 743 was shown to be approximately equal in 5'-TGG and ~50% more efficient in 5'-CGG and 5'-GGC. In general, ET 743 alkylated better those oligonucleotides in which Y was G or C whereas those possessing A or T on both sides of the central guanine were not detectably recognized by the drug even at the higher concentration tested. A drawback of this study, later addressed by Seaman and Hurley [14], was that several of the 16 sample sequences studied actually contained more than one potential binding site for ET 743. This was so either because triplet target sites were overlapped or due to the simultaneous presence of one or more alternative binding sites on the complementary strand of the duplex. When the previous experimental results were reinterpreted on a structural basis [14] a much clearer picture arose and a set of reactivity rules were proposed that also

took into account the hydrogen bonding possibilities. Thus, it was established that in the target sequence, 5'-XGY, the favoured base for Y is either G or C, in agreement with the previous proposal. In addition, a pyrimidine base (T or C) is preferred for X when Y is G whereas a purine (A or G) is needed if Y is C. If both 3' and 5' requirements are met, the sequence is highly reactive, e.g. 5'-AGC, 5'-GGC, 5'-CGG, and 5'-TGG (Fig. 4). Failure to satisfy the first requirement (that Y be G or C) results in low reactivity sequences while fulfillment of only this condition is characteristic of moderately reactive sequences (e.g. 5'-GGG and 5'-AGG). The success of this relatively simple set of rules further supports the idea that reactivity to ET 743 (and ecteinascidins in general) is governed by a combined direct read-out of a central guanine plus two additional flanking bases. This view is amply substantiated by experimental and modelling structural work demonstrating that the extent and spatial orientation of a hydrogen-bonding network involving specific DNA base donors/acceptors does indeed direct the course of sequence recognition.

### Structural Studies of Ecteinascidin-DNA Complexes

Knowledge of the above rules and potential problems is important in order to design a suitable duplex oligonucleotide for structural studies. If the sequence contains, for example, a 5'-AGC target triplet site on one strand followed by a cytosine there will be an alternative binding site, 5'-GGC, on the complementary strand, and this may compromise the success of the experiment. Although inosine could be used to replace guanine at some key positions, as in the band shift experiments reported above, to our knowledge this strategy has not been used yet in NMR studies of DNA-ET complexes.

The first report of a 1:1 ET 743-DNA adduct involved the dodecanucleotide  $d(\text{CGTAA}\underline{\text{GCTTACG}})_2$  [13]. The two-

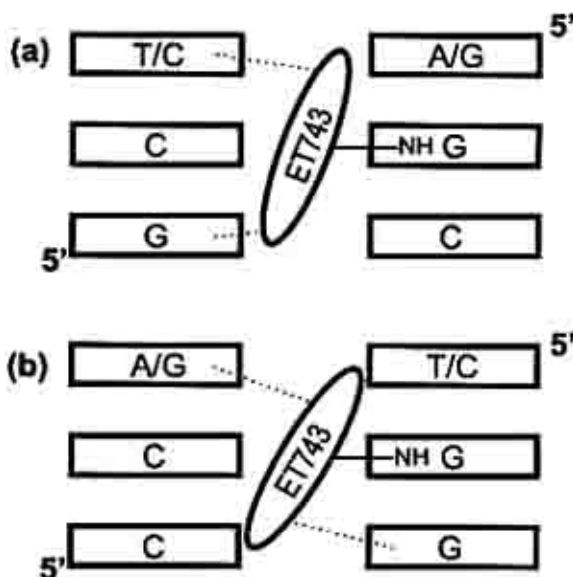


Fig. (4). The two types of sequences favored by ET 743 representing two distinct optimal binding modes [14].

dimensional NMR spectra exhibited well-resolved cross-peaks that could be assigned to a single species and allowed the identification of a number of intermolecular contacts. Several critical connectivities from the nuclear Overhauser effect spectroscopy (NOESY) spectrum supported the proposed covalent bond between C21 and N2(G6) and clearly defined the precise orientation of the drug in the minor groove, with the A subunit to the 5'-side of the alkylated guanine (G6) and the B and C subunits to the 3'-side. Evidence for the interaction of ET 743 with the alkylated DNA strand was provided by nuclear Overhauser effects (NOEs) of the H21 proton into H1'(G6), H1'(C7), H1'(T8), H2'/H2''(C7), H6(C7), and H6(T8) and the H22 protons into H1'(C7) and H3'(T8). Association of subunits A and B with the opposite strand was shown by NOE connectivities between 6Me, 5OAc, H23, H4, and H11 protons of ET 743 to G18, C19, T20, and T21 protons, as well as by aromatic shielding of subunit B over the deoxyribose of G18 and C19. On the contrary, the broad intramolecular NOEs of subunit C suggested interconversion between different conformers and lack of specific association with a site on the DNA, which was supported by the presence of just a single weak ET 743-DNA NOE involving this subunit (H3'(T8)-7'OMe). Taken together, these observations provided evidence that this part of the molecule is perpendicularly projected above the minor groove, in accordance with earlier proposals. One of the major differences between the resulting NMR-based model and the previous theoretical model was that the protonated N12(Me) hydrogen bonds to O2(T20) rather than to the G6:C19 base pair at the alkylation site [11]. Since another hydrogen bond exists between the amino group of G18 and one of the dioxymethylene oxygens, subunit B penetrates more deeply into the minor groove and gets into closer contact with the wall of the nonalkylated strand. Subunit A remains perpendicular to the minor groove and a hydrogen bond is proposed between 18OH and O1'(T20).

On the basis of this first NMR-based model, the importance of suitably placed DNA hydrogen bond acceptors to the 5'-side and a hydrogen bond donor to the 3'-side of the alkylation site was highlighted. Perhaps more importantly, the role of subunits A and B in DNA recognition and bonding was rationalized as well as the paucity of subunit C-DNA interactions.

Further insight into the protonation state of ET-DNA adducts was gained when the same dodecamer was isotopically labelled with  $^{13}\text{C}$  and  $^{15}\text{N}$  [32]. The NMR spectra confirmed lack of protonation on either the exocyclic nitrogen of G6 or N2 of the drug whereas N12 appeared protonated. A mechanism for activation was then proposed that takes advantage of the increased strength of the hydrogen bond between the proton on N12 and the hydroxyl group on C21 as the ET 743 molecule approaches the minor groove and is desolvated. This proton, which is essential for both sequence recognition and adduct stabilization, would then catalyze the dehydration of the carbinolamine yielding the reactive iminium intermediate that undergoes nucleophilic attack at C21 by the exocyclic amino group of G6. The resulting alkylated guanine would be initially protonated but it would regain neutrality by transferring the proton to either N12 of ET 743 or the transiently bound

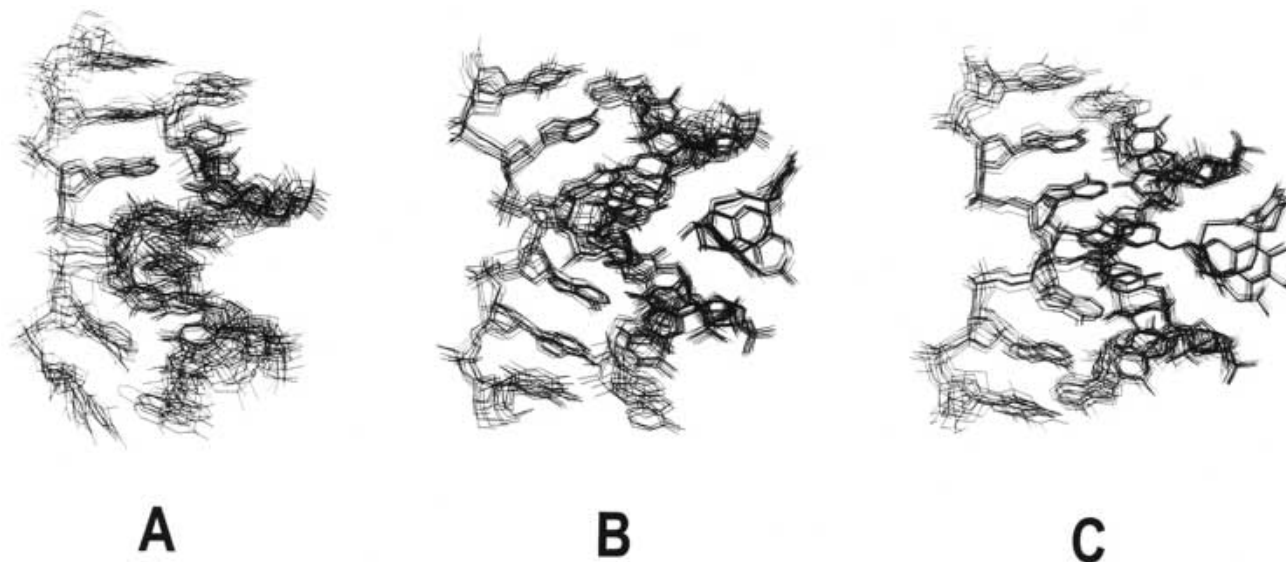
water molecule originated in the previous step. This proton shuttle would restore the original protonation state of the drug and would expel the water molecule as the transition state collapses to the final covalent adduct [14,32]. Since a similar mechanism operates in the activation of pyrrolo(1,4)benzodiazepine antibiotics [10,33], it appears that Nature ensures the reactivity of these carbinolamine-containing molecules by the inclusion of an internal catalytic proton adjacent to the leaving hydroxyl group.

The two-dimensional  $^1\text{H}$  NMR study with the  $\text{d}(\text{CGTAAGCTTACG})_2$  dodecanucleotide was expanded by the same group to include the closely related ET 736, which allowed a better characterization of the hydrogen-bonding network associating the covalently bonded drug with the DNA duplex than in the corresponding adduct with ET 743 [14]. Nevertheless, similar hydrogen bonding interactions for the common A-B-subunit scaffold were deduced from comparison of exchangeable and non-exchangeable  $^1\text{H}$  NMR data. In both cases, unbroken  $^1\text{H}$ - $^1\text{H}$  NOESY cross-connectivity patterns did not reveal any radical structural perturbation in the DNA molecule as a consequence of drug binding. The single set of NMR signals of the unbound self-complementary duplex was of course split into two distinct sets of signals corresponding to the unmodified (C13-G24) and covalently modified (C1-G12) strands.

The limited set of intermolecular distance restraints obtained for this covalent complex was then used in our laboratory to model the precovalent complex and the covalent adduct between ET 743 and the non-self-complementary nonamer  $\text{d}(\text{TAAAGCTTA})_2$ . The interaction was studied by means of unrestrained molecular dynamics (MD) simulations in explicit solvent and in the presence of neutralizing sodium ions, and the complexes were compared with parallel simulations of the free oligos [34]. The prealkylation binding complex showed the protonated N12

hydrogen bonded to the O2 acceptor atom of T15 and the carbinol OH acting both as a hydrogen bond donor to O2 of C14 and as a hydrogen bond acceptor from N2 of G5. A distance (which cannot be measured experimentally) of about 4.0 Å between C21 and N2(G5) suggested that in this complex C21 is a suitable target for nucleophilic attack by G5. Further anchoring to the DNA minor groove was provided by a hydrogen bond between the methylenedioxy oxygen and N2 of G13, as suggested. The OH on subunit C was found to hydrogen bond alternatively to its neighboring methoxy group and to O1P of the phosphate linking T7 and T8 whereas the OH on subunit A was not involved in intermolecular hydrogen bonding. As a consequence of these interactions, the inter-strand O6(G5)-N4(C14) and N1(G5)-N3(C14) hydrogen bonds were broken and both the opening and the negative propeller twist of the G5-C14 base pair increased relative to the free oligonucleotide. This distortion, however, could be easily accommodated by the neighboring base pairs so that the double helical structure was only minimally perturbed except for widening of the minor groove and a net bending towards the major groove. Interestingly, when the precovalent complex of ET 743 with the sequence  $\text{d}(\text{TAACGATTA})_2$  was simulated under the same conditions, the MD trajectory was unstable and no suitable geometry for adduct formation was found, in agreement with the fact that a CGA triplet does not represent a good binding site for the drug [14,31]. Therefore, the simultaneous occurrence of three hydrogen bonds seems to be critical for stabilization of ET-DNA prealkylation complexes, and there are two optimal ways to achieve this, as shown in (Fig. 4).

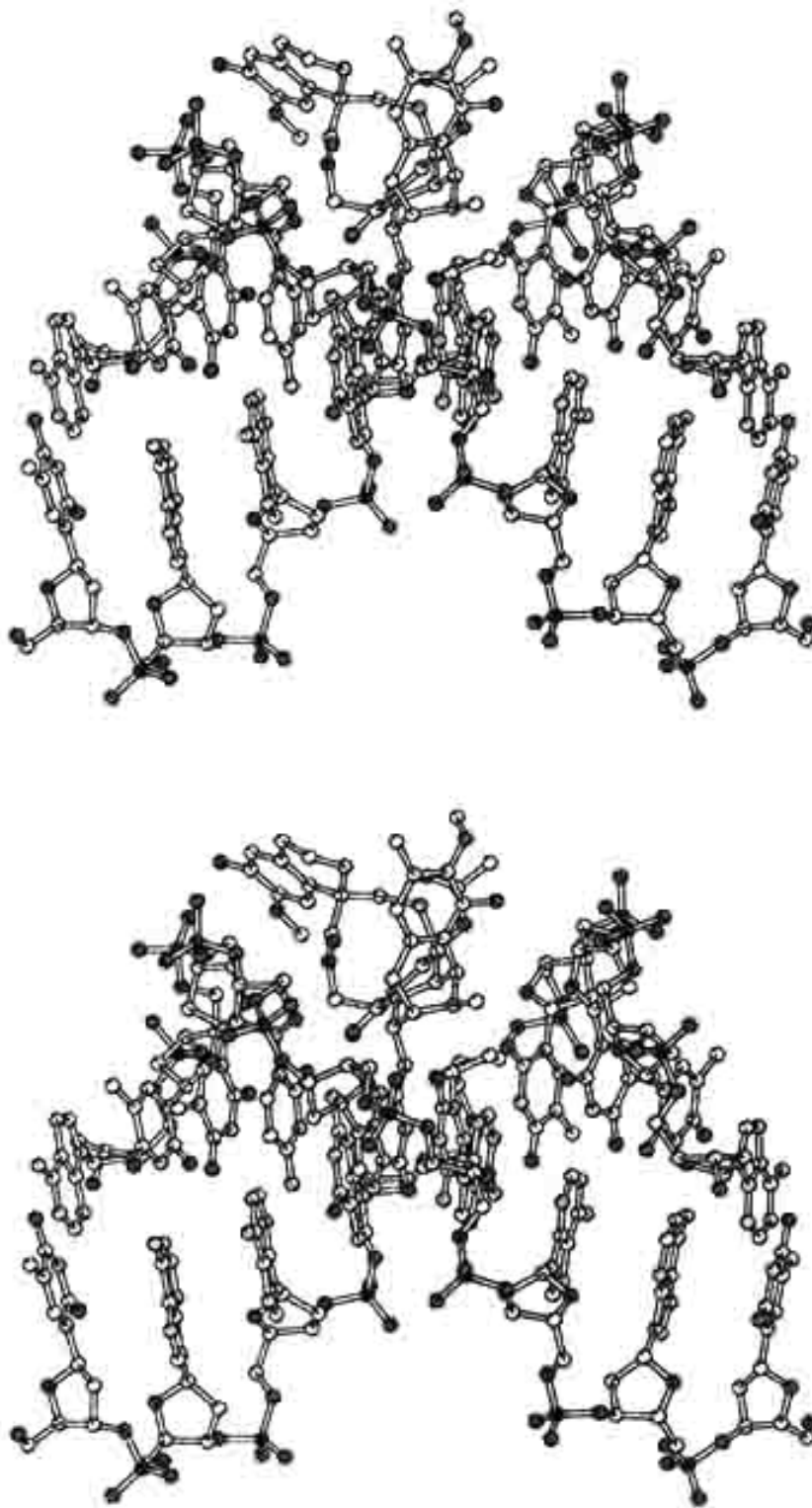
Simulation of the ET 743- $\text{d}(\text{TAAAGCTTA})_2$  and ET 743- $\text{d}(\text{TAACGGTTA})_2$  covalent adducts revealed a considerably enlarged minor groove and significant bending toward the major groove (Fig. 5), mostly due to an increase in positive roll at the base pair step involved in covalent bond



**Fig. (5).** View of the minor groove in the oligo  $\text{d}(\text{TAAAGCTTA})_2$  free (A), in the pre-alkylation complex with ET 743 (B), and in the covalent adduct (C). The 10 superimposed models shown in each case correspond to consecutive 50-ps average structures from an unrestrained molecular dynamics simulation in aqueous solution [34]. Average values of the width of the minor groove in the central part of each duplex, measured as the shortest inter-phosphate distances across the groove (P-P distance minus 5.8 Å), are 10.5, 14.2 and 14.1 Å (C6-T16) and 10.7, 14.6 and 14.9 Å (T7-T15), respectively.

formation. Similar results were obtained for the related Pt 650 bonded to d(TAACGGTTA)<sub>2</sub>. Interestingly, the central CGG triplet appeared to accommodate both ET 743 and Pt 650 with less distortion than AGC, suggesting that the intrinsic bendability of a particular DNA sequence and its overall preorganization could facilitate specific recognition

by ET 743. Coordinates for representative refined average structures from the MD simulations of the ET 743-d(TAAAGCTTA)<sub>2</sub> (Fig. 6) and ET 743-d(TAACGGTTA)<sub>2</sub> complexes are available from the Research Collaboratory for Structural Bioinformatics with PDB identification codes 1EZ5 and 1EZH, respectively.



**Fig. (6).** MolScript [77] stereoview of a representative covalent adduct formed between ET 743 and the nonanucleotide d(TAAAGCTTA)<sub>2</sub>.

The validity of the results from these MD simulations was supported by evidence obtained previously from gel migration and cyclization kinetics experiments carried out with a 21-bp oligonucleotide containing a single guanine in the middle of a favoured (5'-AGC) triplet [35]. The 3' cytosine was paired to an inosine on the noncovalently modified strand and a 3-base overhang was included to allow for head-to-tail self-ligation in the presence of T4 DNA ligase. The ligated 21-mers containing the ET 743 adduct showed a retardation in electrophoretic mobility relative to the unmodified linear multimers, and a greater tendency to form closed circular DNA. Both observations were suggestive of induced bending, and the angle of absolute curvature was estimated to be  $17 \pm 3^\circ$ , in good agreement with our estimates from the MD simulations [34]. To determine the directionality of the induced bend a series of four oligonucleotides was designed in which the same single alkylation site as before was positioned 5, 7, 11, and 13 base pairs (bp) away from the center of a tract containing five consecutive adenines. Despite the fact that X-ray crystallography has consistently shown them as straight [36], these A-tracts are known to bend DNA toward the minor groove and this bend can be used as a reference point for studying curvature by placing the A-tract "in phase" or "out of phase" with another bending element [37]. Since the anomalous mobility was at a maximum when the ET 743 alkylation site was positioned 5 bp (one-half helical turn) away from the centre of the A-tract and at a minimum when the separation was 11 bp (a complete helical turn), these observations were taken as an indication that the bend caused by ET 743 was in a direction opposite to that of the A-tract, that is, toward the major groove, again in good correspondence with the findings from our MD simulations. This feature was novel among minor groove DNA monoalkylating agents as covalent modification of N3 of adenine in AT-rich regions by (+)-CC-1065 and related compounds is accompanied by bending of the DNA into the minor groove [38].

### Possible Biomolecular Targets of Ecteinascidins

The mechanism of action of ET 743 and related compounds is still the matter of intense investigation in several laboratories. In view of its profile in the National Cancer Institute 60 Cell Line Human Tumor Screen and its outstanding potency, ET 743 is thought to mediate cell killing through a mechanism different from that employed by other clinically used drugs [17,39,40]. Although the accumulating evidence suggests that the cytotoxic effect of ET 743 is exerted by interaction with DNA, the details of the mechanism operating beyond simple DNA alkylation remain obscure.

Early bioassays using purified ecteinascidins demonstrated inhibition of RNA polymerase activity and decreased RNA and DNA synthesis [3,4,12,15]. More recently, exposure of [<sup>3</sup>H]thymidine-prelabelled cells to ET 743 and Pt 650 for 1 h, followed by lysis with sodium dodecyl sulfate and treatment with potassium chloride, showed the presence of radioactivity in the protein pellet, indicative of drug-induced DNA protein cross-linking. By

using melanoma (A375) and lung (A-549) tumour cells, the magnitude of cross-linking was shown to be greater in the more sensitive cell line (A375). A subsequent immunoblot assay carried out to evaluate whether this reaction was dependent on topoisomerase determined that the cross-linking ability of ET 743 and Pt 650 was not dependent on topoisomerase II but could be mediated by topoisomerase I (topo I) [17]. The finding that DNA minor groove alkylation by ET 743 induces topo I-mediated protein-linked DNA breaks *in vitro* and *in vivo* was later confirmed by other authors [39] in what represents the first report of the entrapment of this enzyme by a DNA-alkylating drug. Nevertheless, the observation that optimum DNA-topo I cross-linking is achieved at micromolar concentrations, together with the fact that ET 743 and Pt 650 show the same cytotoxic effects on cells sensitive and resistant to the topoisomerase inhibitors etoposide and camptothecin, appear to suggest that topoisomerase activity may be only an auxiliary effect incidental to the primary mechanism of action of these unique agents [17]. The same conclusion applies to the report that, also at relatively high concentrations, ET 743 disorganises the microtubule network in the absence of interaction with tubulin [41].

The reported ET 743-induced DNA bending [34,35] and DNA minor groove widening [14,34] could be important for either impairing the binding of sequence-specific transcription factors [42] or enhancing recognition by DNA binding proteins such as those that recognize cis-platinated DNA [43]. Disruption of protein-DNA association could be brought about by drug-induced bends incompatible with the direction of bending required by the DNA-binding protein at its recognition site. Alternatively, the existing bend could be advantageous for binding of some other proteins to the major groove, but no such proteins have been identified so far. On the other hand, if multiple potential binding sites for ET 743 are properly phased in a relatively short stretch of DNA, the imposed cumulative curvature could bring closer together specified fragments not contiguous in primary sequence [44] and the drug could be serving a surrogate protein function. This has been demonstrated for (+)-CC-1065, which binds to the DNA minor groove and can partially substitute for the transcription factor Sp1 in up-regulating the level of basal gene expression [45].

Since there are many occurrences of the triplet sequences favoured by ET 743 and related compounds in genomic DNA, one may also wonder whether the cytotoxic effect caused at very low concentrations (between pM and nM, depending on cell type) relies on binding to a specific DNA sequence or to a particular DNA architectural motif. A sequence that has elicited some interest is the TGG triplet contained in the complementary strand of the so-called CCAAT box, a common element in eukaryotic promoters that can be found in both the forward and the reverse orientation (*i.e.* ATTGG). These five nucleotides are absolutely required for binding of the evolutionary conserved activator NF-Y (CBF, HAP2/3/4/5) [46] to the DNA minor groove and, in fact, binding of NF-Y to this sequence has been shown to be inhibited by ET 743 [47]. While it seems clear that occupancy of the minor groove by the covalently bonded drug will preclude association of NF-Y with its cognate DNA element, the micromolar concentrations

necessary to achieve this effect, together with the fact that stably bound NF- $\kappa$ B is not displaced from the CCAAT box by the low concentrations of ET 743 used both in cell cultures and in patients [48], point to a different DNA site as the likely target for ET 743.

More recently, the sensitivity to ET 743 was shown to be much higher in G<sub>1</sub>-synchronised cells than in cells synchronised in either S phase or the G<sub>2</sub>/M interphase [40]. In addition, p53 levels were significantly increased upon ET 743 treatment in cell lines expressing a wild-type phenotype although p53(-/-) cells were equally sensitive to the drug. On the other hand, cells deficient in nucleotide excision repair were found to be more resistant to the cytotoxic effects of ET 743 than control cells [40], in agreement with reports suggesting a dependency on proteins involved in transcription-coupled repair for the antiproliferative action of ET 743 [49].

Other possibilities notwithstanding, we recently noted a remarkable structural parallel in the manner ET 743 and the zinc finger-containing DNA binding domain of the early growth response protein 1 (EGR1) and other transcription factors induce DNA distortions [50]. This observation led us to propose that binding of ET 743 to a binary complex made up of a suitable DNA GC-rich region to which a EGR1 or a Sp1-like zinc finger-containing transcription factor is bound should be favoured over binding to naked DNA due to the protein-induced DNA structural preorganization [51]. Interestingly, Sp1-like proteins regulate the expression of more than 1000 different genes involved in cell proliferation, differentiation and apoptosis through their binding to GC-rich *cis*-regulatory sequences and subsequent interaction with the basal transcription machinery [52,53]. These sequences are present in the promoters of genes encoding, for example, cell cycle regulators (p15/INK4B, p21/WAF1, p27/KIP1), MAP kinases, regulatory GTPases (Ha-Ras), histones, enzymes involved in DNA synthesis (thymidine kinase, dihydrofolate reductase), growth factors (TGF- $\beta$ 1, PDGF), and growth factor receptors (insulin receptor, insulin-like growth receptor). The transcription factors which bind to these promoter sequences are classified into different subfamilies (e.g. Sp, BTEB, KLF, CPBP, and TIEG) and can behave as either activators or repressors. A characteristic of this family, also shared by EGR1, is the presence of three Cys<sub>2</sub>His<sub>2</sub>-type zinc fingers in their C-terminal regions that “read” the major groove and confer sequence-specific DNA binding properties [54]. Even though some anticancer drugs (e.g. nogalamycin and chromomycin A<sub>3</sub>) are known to interfere with binding of some of these transcription factors to their respective recognition sequences [42], in the case of ET 743, on the contrary, concomitant binding of the drug and the protein to the same DNA stretch is possible as protein and drug occupy opposite faces of the DNA by filling either the major or the minor groove, respectively.

EGR1 and the ubiquitously expressed Sp1 recognize different, but related, nucleotide sequences that can be partially overlapped [55], in which case their binding is mutually exclusive. Binding of one or the other protein will depend on their intrinsic affinities, which can be finely tuned by phosphorylation [56,57], and their relative concentrations. We hypothesize that the resulting balance in such

competition can be influenced by the presence of bonded ET 743 to one or more sites. Some DNA triplet subsites, such as TGG or GGC, can provide simultaneous binding sites for a zinc finger (in the major groove) and ET 743 (in the minor groove). In other instances, the ET 743 binding site may overlap two adjacent zinc finger subsites, and the possibility also exists that, given the right sequence, two or more ET 743 molecules are bonded consecutively. One example could be the CGGGCCCGG sequence present in the *p21/WAF1* promoter, where underlined bases indicate ET 743 binding sites and dots mean binding to the complementary strand. This sequence comprises part of a Sp1 binding site located at -69 relative to the transcription start site that mediates p53-independent activation of the *p21/WAF1* promoter by the histone deacetylase inhibitors trichostatin A [58] and suberoylanilide hydroxamic acid (SAHA) [59]. This observation could be interesting as SAHA has been shown to induce cell differentiation of cultured murine erythroleukemia cells and inhibit growth in human breast cancer cells [60].

Sp1 bound to GC-rich positive regulatory elements is known to be displaced in a dose-dependent manner by elevated amounts of EGR1 [61], which results in decreased Sp1-dependent transcription. Moreover, *EGR1* gene expression is down-regulated by EGR1 itself and activated by Sp1 [62]. It is therefore likely that the interplay between EGR1, Sp1 and similar transcription factors is affected by the presence of ET 743 covalently bonded to one or more of their recognition sites.

The chromosomal location of EGR1 is 5q23-31, and this region is often deleted in epithelial hyperproliferation and myeloid disorders [63,64]. For example, human HT-1080 fibrosarcoma cells, which happen to be particularly sensitive to the antitumour effects of ET 743, are known to express little or no EGR1, and only mutant p53. In these cells, stable expression of exogenous EGR1 has been shown to inhibit transformed growth in a dose-dependent manner and to cause decreased tumorigenicity [65]. Even though growth inhibition in human HT1080 fibrosarcoma cells can also be brought about by overexpression of just the DNA binding domain of this transcription factor [65], this antiproliferative activity has been proposed to arise, at least in part, from the ability of EGR1 to induce the expression of human transforming growth factor  $\beta$ 1 (TGF- $\beta$ 1) and p21/WAF1 [66,67], together with increased secretion of fibronectin and plasminogen activator inhibitor-1, which may contribute to the suppression by causing enhanced cell attachment [68]. The EGR1-stimulated transactivation can be inhibited by expression of the Wilms' tumor suppressor (WT1) [69], a known specific DNA-binding competitor possessing four zinc fingers in its DNA binding domain. WT1 [70] binds to the EGR1 consensus site, as well as other sites, and may repress (e.g. *TGF- $\beta$ 1*) or stimulate (e.g. *bcl-2*) gene expression depending on promoter context. Since DNA binding of WT1, in turn, has been recently shown to be inhibited by association with members of the p53 family of tumour suppressors [71], a complex pattern of cross-regulation and important roles in cell growth, differentiation and programmed cell death are apparent for many of these transcription factors whose activity might be modulated by the presence of DNA-bonded ET 743.

Similar growth suppression effects by EGR1 overexpression have been observed in other cell lines, including mammary (ZR-75-1, MDA-MB-468, T-B20, MCF-7, T-47D) and lung tumours, glioblastoma U251 [72], and osteogenic sarcoma SAOS-2, but not in HeLa or PC-3 prostate carcinoma cells [69]. These findings suggest that EGR1 expression is necessary for normal growth control, but additional evidence indicates that these regulatory effects are cell type-specific. Thus, EGR1 is overexpressed in a majority of human prostate cancers where it is implicated in the regulation of several genes important for prostate tumor progression (e.g. insulin-like growth factor-II, platelet-derived growth factor-A, and TGF- $\beta$ 1) [73]. In fact, by using transgenic mouse models of prostate cancer it has been recently shown that tumor progression from prostatic intra-epithelial neoplasia to invasive carcinoma is significantly impaired in *EGR1*<sup>-/-</sup> mice whereas tumor initiation and tumor growth rate are not affected by the lack of EGR1 [74]. On the other hand, subtypes of smooth muscle cells expressing abundant levels of Sp1 have been shown recently to produce the death agonist, Fas ligand, and undergo greater spontaneous apoptosis [75].

Since ET 743 widens the minor groove and bends DNA toward the major groove, this drug and related analogues can be expected to bind more tightly to locally bent DNA-protein complexes displaying these structural characteristics than to linear B-DNA [51]. DNA sequence-dependent deformability is indeed recognized as an important contributor to specific recognition by proteins and small ligands, and affinity differences among EGR1 family members have been shown to reflect primarily rates of dissociation from the DNA [76]. Thus, we would expect additional stabilization upon binding of ET 743 to a complex formed between a suitable DNA regulatory sequence and a zinc finger-containing protein such as EGR1, WT1, and/or members of the Sp1-like family of transcription factors. The outcome of these interactions is likely to be complex and dependent on cellular type, promoter context and phase of the cell cycle. Since only a subset of the sequences found in these GC-rich regulatory regions provide suitable anchoring sites to ET 743, we believe that this drug, in addition to its promising role as an antitumor agent, will prove to be a very useful tool in dissecting the transcriptional regulation of several gene classes.

## CONCLUSIONS

The high expectations raised in 1969 by the potent antiproliferative activities of natural ecteinascidins have been realized by the optimistic results obtained from phase I and II clinical trials with ET 743 and the demonstration of both a high potency and a favourable therapeutic index. Limitations on testing initially imposed by the low amounts of drug present in the natural source have been overcome by current availability in kilogram quantities from chemical modification of microbially produced cyanosauracin B. ET 743 alkylates DNA on the exocyclic amino group of guanines following noncovalent binding to the minor groove of some preferred triplet sequences in a reaction that can be reversed on DNA denaturation. Sequence selectivity has been rationalized on the basis of a direct read-out through a

set of well-defined hydrogen-bonding rules dictating that the preferred targets are RGC and YGG, where R and Y stand for purine and pyrimidine, respectively, and the underlined guanine is the base that undergoes alkylation by ET 743. Binding of an ET 743 molecule to a DNA target site brings about widening of the minor groove and curvature toward the major groove. The resulting DNA deformation strongly resembles that induced by zinc finger-containing transcription factors, such as EGR1 and Sp1, which bind to GC-rich regulatory sequences in many gene promoters. Since the transcription factor and ET 743 can bind simultaneously on opposite sides of the double helix, we have put forward the hypothesis that binding of ET 743 to the fully accessible minor groove in well-defined steps of Sp1-DNA or EGR1-DNA complexes is facilitated by protein-induced structural changes and could lead to increased complex stabilization. The resulting enhanced specificity could account for the unusually high potency and the remarkable effects on gene transcription of ET 743 and other ecteinascidin-like molecules. Since EGR1 and Sp1 binding sites are frequently overlapped, and both proteins belong to larger families of related transcription factors with important roles in cell growth, differentiation and programmed cell death, the effects of ET 743 binding to some of these sites is not easily predictable but can have far reaching consequences, including the known regulation of abnormal growth and the triggering of apoptosis. We hope the ideas summarized in this review will provide other researchers with a number of places to look for answers to our questions about possible mechanisms.

## ABBREVIATIONS

Alloc	=	Allyloxycarbonyl
Boc	=	<i>t</i> -Butoxycarbonyl
Bp	=	Base pairs
CAN	=	Ceric ammonium nitrate
DBU	=	1,8-Diazabicyclo[5.4.0]undec-7-ene
DDQ	=	2,3-Dichloro-5,6-dicyano-1,4-benzoquinone
EGR1	=	Early growth response protein 1
ET	=	Ecteinascidin
FABMS	=	Fast Atom Bombardment Mass Spectrometry
Fmoc	=	9-Fluorenylmethoxycarbonyl
MD	=	Molecular dynamics
MOM	=	Methoxymethyl
MS	=	Mass spectrometry
NMR	=	Nuclear magnetic resonance
NOE	=	Nuclear Overhauser effect

NOESY = Nuclear Overhauser effect spectroscopy

PDB = Protein Data Bank

SAHA = Suberoylanilide hydroxamic acid

TFA = Trifluoroacetic acid

TGF- 1 = Transforming growth factor 1

TMSI = Trimethylsilyl iodide

TROC = , , , Trichloroethoxycarbonyl

## REFERENCES

- [1] Sigel, M.M.; Wellham, L.L.; Lichter, W.; Dudeck, L.E.; Gargus, J.L.; Lucas, A.H. In *Food-Drugs from the Sea Proceedings*; Youngken, H.W., Ed.; Marine Technology Society: Washington DC, **1969**; pp. 281-294.
- [2] Rinehart, K.L.; Holt, T.G.; Fregeau, N.L.; Stroh, J.G.; Keifer, P.A.; Sun, F.; Li, H.; Martin, D.G. *J. Org. Chem.*, **1990**, *55*, 4508.
- [3] Rinehart, K.L. *Med. Res. Rev.*, **2000**, *20*, 1.
- [4] Sakai, R.; Jares-Erijman, E.; Manzanares, I.; Elipe, M.V.; Rinehart, K.L. *J. Am. Chem. Soc.*, **1996**, *118*, 9017.
- [5] Ikeda, Y.; Idemoto, H.; Hirayama, F.; Yamamoto, K.; Iwao, K.; Asao, T.; Munakata, T. *J. Antibiot. (Tokyo)*, **1983**, *36*, 1279.
- [6] Irschik, H.; Trowitzsch-Kienast, W.; Gerth, K.; Hofle, G.; Reichenbach, H. *J. Antibiot. (Tokyo)*, **1988**, *41*, 993.
- [7] Okumoto, T.; Kawana, M.; Nakamura, I.; Ikeda, Y.; Isagai, K. *J. Antibiot. (Tokyo)*, **1985**, *38*, 767.
- [8] Ishiguro, K.; Takahashi, K.; Yazawa, K.; Sakiyama, S.; Arai, T. *J. Biol. Chem.*, **1981**, *256*, 2162.
- [9] Kohn, K.W.; Glaubiger, D.; Spears, C.L. *Biochim. Biophys. Acta*, **1974**, *361*, 288.
- [10] Lown, J.W.; Joshua, A.V.; Lee, J.S. *Biochemistry*, **1982**, *21*, 419.
- [11] Guan, Y.; Sakai, R.; Rinehart, K. L.; Wang, A. H. *J. Biomol. Struct. Dyn.*, **1993**, *10*, 793.
- [12] Sakai, R.; Rinehart, K.L.; Guan, Y.; Wang, H.J. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 11456.
- [13] Moore, B. M. II; Seaman, F. C.; Hurley, L. H. *J. Am. Chem. Soc.*, **1997**, *119*, 5475.
- [14] Seaman, F. C.; Hurley, L. H. *J. Am. Chem. Soc.*, **1998**, *120*, 13028.
- [15] Morales, J.J. Ph. D. Dissertation, University of Illinois, Urbana, **1999**.
- [16] Lubber-Narod, J.; Smith, B.; Grant, W.; Jimeno, J.; López-Lázaro, L.; Faircloth, G. AACR, 92<sup>nd</sup> Annual Meeting. New Orleans, 24-28 March, **2001**. Abstract number 1133.
- [17] Martinez, E. J.; Owa, T.; Schreiber, S. L.; Corey, E. J. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 3496.
- [18] Fukuyama, T.; Sachleben, R.A. *J. Am. Chem. Soc.*, **1982**, *104*, 4957.
- [19] Kubo, A.; Saito, N.; Nakamura, M.; Ogata, K.; Sakai, S. *Heterocycles*, **1987**, *26*, 1765. (b) Kubo, A.; Saito, N.; Yamauchi, R.; Sakai, S. *Chem. Pharm. Bull.*, **1987**, *35*, 2158. (c) Kubo, A.; Saito, N.; Yamato, H.; Masubuchi, K.; Nakamura, M. *J. Org. Chem.*, **1988**, *53*, 4295.
- [20] Kubo, A.; Saito, N.; Yamato, H.; Yamauchi, R.; Hiruma, K.; Inoue, S. *Chem Pharm. Bull.*, **1988**, *36*, 2607. (b) Saito, N.; Yamauchi, R.; Nishioka, H.; Ida, S.; Kubo, A. *J. Org. Chem.*, **1989**, *54*, 5391.
- [21] Saito, N.; Ohira, Y.; Wada, N.; Kubo, A. *Tetrahedron*, **1990**, *46*, 7711.
- [22] Saito, N.; Nishida, M.; Kubo, A. *Chem. Pharm. Bull.*, **1991**, *39*, 1343.
- [23] Fukuyama, T.; Yang, L.; Ajeck, K. L.; Sachleben, R. A. *J. Am. Chem. Soc.*, **1990**, *112*, 3712.
- [24] Fukuyama, T.; Linton, S.D.; Tun, M.M. *Tetrahedron Lett.*, **1990**, *31*, 5989.
- [25] Myers, A. G.; Kung, D. W.; *J. Am. Chem. Soc.*, **1999**, *121*, 10828.
- [26] Myers, A. G.; Kung, D. W. *Org. Lett.*, **2000**, *2*, 3019.
- [27] Martinez, E. J.; Corey, E. J. *Org. Lett.*, **1999**, *1*, 75.
- [28] Zhou, B.; Edmondson, S.; Danishefsky, S. J. *Tetrahedron Lett.*, **2000**, *41*, 2039. (b) Zhou, B.; Danishefsky, S. J. *Tetrahedron Lett.*, **2000**, *41*, 2043.
- [29] Corey, E. J.; Gin, D. Y.; Kania, R. *J. Am. Chem. Soc.*, **1996**, *118*, 9202.
- [30] Cuevas, C.; Pérez, M.; Martín, M. J.; Chicharro, J. L.; Fernández-Rivas, C.; Flores, M.; Francesch, A.; Gallego, P.; Zarzuelo, M.; De la Calle, F.; García, J.; Polanco, C.; Rodríguez, I.; Manzanares, I. *Org. Lett.*, **2000**, *2*, 2545.
- [31] Pommier, Y.; Kohlhagen, G.; Bailly, C.; Waring, M.; Mazumder, A.; Kohn, K. W. *Biochemistry*, **1996**, *35*, 13303.
- [32] Moore, B. M., II; Seaman, F. C.; Wheelhouse, R. T.; Hurley, L. H. *J. Am. Chem. Soc.*, **1998**, *120*, 2490. (b) *ibid.* (corrigenda) *J. Am. Chem. Soc.*, **1998**, *120*, 9975.
- [33] Hurley, L.H.; Gairola, C.; Zmijewski, M. *Biochim. Biophys. Acta*, **1977**, *475*, 521.
- [34] García-Nieto, R.; Manzanares, I.; Cuevas, C.; Gago, F. *J. Am. Chem. Soc.*, **2000**, *122*, 7172.
- [35] Zewail-Foote, M.; Hurley, L. H. *J. Med. Chem.*, **1999**, *42*, 2493.
- [36] Goodsell, D.S.; Dickerson, R.E. *Nucleic Acids Res.*, **1994**, *22*, 5497.
- [37] Zinkel, S.S.; Crothers, D.M. *Nature*, **1987**, *328*, 178.
- [38] Lee, C.S.; Sun, D.; Kizu, R.; Hurley, L.H. *Chem. Res. Toxicol.*, **1991**, *4*, 203.
- [39] Takebayashi, Y.; Pourquier, P.; Yoshida, A.; Kohlhagen, G.; Pommier, Y. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 7196.

- [40] Erba, E.; Bergamaschi, D.; Bassano, L.; Damia, G.; Ronzoni, S.; Faircloth, G.T.; D'Incalci, M. *Eur. J. Cancer*, **2001**, 37, 97.
- [41] García-Rocha, M.; García-Gravalos, M. D.; Avila, J. *Br. J. Cancer*, **1996**, 73, 875.
- [42] Welch, J.J.; Rauscher, F.J.; Beerman, T.A. *J. Biol. Chem.*, **1994**, 269, 31051.
- [43] Ohndorf, U. M.; Rould, M. A.; He, Q.; Pabo, C. O.; Lippard, S. J. *Nature*, **1999**, 399, 708.
- [44] Zewail-Foote M., Hurley, L. H. *Anticancer Drug Des.*, **1999**, 14, 1.
- [45] Sun, D.; Hurley, L.H. *Biochemistry*, **1994**, 33, 9578.
- [46] Mantovani, R. *Nucleic Acids Res.*, **1998**, 26, 1135.
- [47] Bonfanti, M.; La Valle, E.; Fernandez-Sousa Faro, J. M.; Faircloth, G.; Caretti, G.; Mantovani, R.; D'Incalci M. *Anticancer Drug Des.*, **1999**, 14, 179.
- [48] Jin, S.; Gorfajn, B.; Faircloth, G.; Scotto, K. W. *Proc. Natl. Acad. Sci. USA*, **2000**, 97, 6775.
- [49] Pourquier, P.; Emmert, S.; Ueda, T.; Takebayashi, Y.; Kraemer, K.H.; Pommier, Y. AACR, 92<sup>nd</sup> Annual Meeting. New Orleans, 24-28 March, **2001**. Abstract number 2987.
- [50] Nekludova, L.; Pabo, CO. *Proc. Natl. Acad. Sci. USA*, **1994**, 91, 6948.
- [51] García-Nieto, R.; Manzanares, I.; Cuevas, C.; Gago, F. J. *Med. Chem.*, **2000**, 43, 4367.
- [52] Philipsen, S.; Suske, G. *Nucleic Acids Res.*, **1999**, 27, 2991.
- [53] Cook, T.; Gebelein, B.; Urrutia, R. *Ann. N. Y. Acad. Sci.* **1999**, 880, 94.
- [54] Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, 29, 183.
- [55] Khachigian, L.M.; Lindner, V.; Williams, A.J.; Collins, T. *Science*, **1996**, 271, 1427.
- [56] Cao, X.; Mahendran, R.; Guy, G.R.; Tan, Y.H. *J. Biol. Chem.*, **1992**, 267, 12991.
- [57] Huang, R.P.; Fan, Y.; deBelle, I.; Ni, Z.; Matheny, W.; Adamson, E.D. *Cell Death Differ.*, **1998**, 5, 96.
- [58] Sowa, Y.; Orita, T.; Hiranabe-Minamikawa, S.; Nakano, K.; Mizuno, T.; Nomura, H.; Sakai, T. *Ann. N. Y. Acad. Sci.*, **1999**, 886, 195.
- [59] Huang, L.; Sowa, Y.; Sakai, T.; Pardee, A.B. *Oncogene*, **2000**, 19, 5712.
- [60] Huang, L.; Pardee, A.B. *Mol. Med.*, **2000**, 6, 849.
- [61] Liu, C.; Calogero, A.; Ragona, G.; Adamson, E.; Mercola, D. *Crit. Rev. Oncol.*, **1996**, 7, 101.
- [62] Cao, X.; Mahendran, R.; Guy, G.R.; Tan, Y.H. *J. Biol. Chem.*, **1993**, 268, 16949.
- [63] Le Beau, M. M.; Espinosa, R., III; Neuman, W. L.; Stock, W.; Roulston, D.; Larson, R. A.; Keinanen, M.; Westbrook, C. A. *Proc. Nat. Acad. Sci. USA*, **1993**, 90, 5484.
- [64] Horrigan, S.K.; Arbieva, Z.H.; Xie, H.Y.; Kravarusic, J.; Fulton, N.C.; Naik, H.; Le, T.T.; Westbrook, C.A. *Blood*, **2000**, 95, 2372.
- [65] Huang, R.-P.; Liu, C.; Fan, Y.; Mercola, D.; Adamson, E. *Cancer Res.*, **1995**, 55, 5054.
- [66] de Belle, I.; Huang, R.P.; Fan, Y.; Liu, C.; Mercola, D.; Adamson, E.D. *Oncogene*, **1999**, 18, 3633.
- [67] Liu, C.; Rangnekar, V.M.; Adamson, E.; Mercola, D. *Cancer Gene Ther.*, **1998**, 5, 3.
- [68] Liu, C.; Yao, J.; de Belle, I.; Huang, R.P.; Adamson, E.; Mercola, D. *J. Biol. Chem.*, **1999**, 274, 4400.
- [69] Liu, C., Adamson, E., Mercola, D. *Proc. Natl. Acad. Sci. USA*, **1996**, 93, 11831.
- [70] Menke, A.L.; van der Eb, A.J.; Jochemsen, A.G. *Int. Rev. Cytol.*, **1998**, 181, 151.
- [71] Scharnhorst, V.; Dekker, P.; van der Eb, A.J.; Jochemsen, A.G. *J. Biol. Chem.*, **2000**, 275, 10202.
- [72] Liu, C.; Yao, J.; Mercola, D.; Adamson, E. *J. Biol. Chem.*, **2000**, 275, 20315.
- [73] Svaren, J.; Ehrig, T.; Abdulkadir, S.A.; Ehrenguber, M.U.; Watson, M.A.; Milbrandt, J. *J. Biol. Chem.*, **2000**, 275, 38524.
- [74] Abdulkadir, S.A.; Qu, Z.; Garabedian, E.; Song, S.K.; Peters, T.J.; Svaren, J.; Carbone, J.M.; Naughton, C.K.; Catalona, W.J.; Ackerman, J.J.; Gordon, J.I.; Humphrey, P.A.; Milbrandt, J. *Nature Med.*, **2001**, 7, 101.
- [75] Kavurma, M.M.; Santiago, F.S.; Bonfoco, E.; Khachigian, L.M. *J. Biol. Chem.*, **2000**, 276, 4964.
- [76] Swirnoff, A.H.; Milbrandt, J. *Mol. Cell. Biol.*, **1995**, 15, 2275.
- [77] Kraulis, P. J. *J. Appl. Crystall.*, **1991**, 24, 946.