Similar Interactions of Natural Products with Biosynthetic Enzymes and Therapeutic Targets explain why Nature produces such a Large Proportion of Existing Drugs

Esther Kellenberger[a], Andreas Hofmann[b], and Ronald J. Quinn*[ab]

Natural products are made by nature through interacting with biosynthetic enzymes. Natural products also exert their effect as drugs by interaction with proteins. Does the recognition of the natural product by biosynthetic enzymes translate to recognition of the therapeutic target? Molecular modelling of flavonoid biosynthetic enzymes and protein kinases with a series of natural product kinase inhibitors led to the development of the concept of Protein Fold Topology (PFT). PFT describes cavity recognition points unrelated to protein fold similarity. The topology or spatial properties are preserved even though there is deformation of the protein elements that participate in the protein-ligand interactions. We observe helices or β-sheets as equivalent in providing the invariant topology for protein-ligand interaction. In this review, we explore: Will PFT aid Drug Discovery or Is it the Reason Natural Products have Drug Properties?

1. Introduction

Nature has proven to be one of the richest sources of therapeutically useful compounds. Natural products are produced biosynthetically within organisms, in order to aid their survival. These compounds may function within the organism itself or externally. In the latter case, natural products act either as toxins to inhibit a critical pathway within other organisms, or as pheromones in order to attract or repel or other organisms.

This highlight provides a perspective of the significant impact natural products have made on the discovery of therapeutic agents by linking interactions of these compounds with the biosynthetic protein on the one hand, and the target/receptor protein on the other. The intriguing question is thus whether such interactions during biosynthesis can be leveraged as a drug discovery tool.

Drug discovery requires the identification of molecules that interact with biological systems. The relevant biological structure space comprises the protein binding sites for potential ligands. Ligands are recognized by proteins in specific binding pockets, complementary in shape and physicochemical properties to the accommodated small molecule. If two binding pockets of different proteins share a common motif, it is likely that ligands or ligand fragments that bind within one binding pocket will also be recognized in the respective part of the other binding pocket.

Ligand binding sites in disease targets serve the purpose of binding the endogenous ligand. Firstly, modulation of a disease target by a drug compound is more likely, if the drug and the endogenous ligand share common interactions with the target protein. These common interactions would be a consequence of a shared complementarity between biological and chemical structure space between the endogenous ligand and the drug.

Secondly, since natural products are of biosynthetic origin, they possess an imprint of the biological structure space in the biosynthetic enzyme. The endogenous ligand therefore carries information about two different binding sites on proteins, the biosynthetic enzyme and the target protein.

The drug discovery task is to find compounds that are complementary to this structure space by correlating biological structure space with chemical structure space (Figure 1). However, the chemical structure space as produced in synthetic libraries often fails to overlap sufficiently with biological structure space.
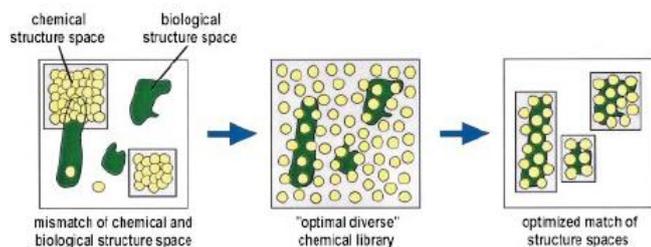
Fig. 1 Chemical diversity as produced in chemical libraries often fails to interact with biological structure space [Box 1]. If the entire chemical structure space could be synthesizes, the HTS would automatically identify suitable molecules. [Box2] As it is impossible to populate all chemical diversity the matching of biological and chemical structure spaces is required to optimise the process of drug discovery. .Reproduced with permission. Wess, G., Urmann, M., Sickenberger, B. Medicinal Chemistry: Challenges and Opportunities. Angew. Chem., Int. Ed. Engl., 2001, 40(18), 3341-3350. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.

If we can unravel the common motif shared by the binding pockets of biosynthetic enzymes and the binding pockets of therapeutic targets, then we would not only understand the reason for the higher relevance of natural products in the development of therapeutics, but also obtain a tool improving rationale probing of biological space to obtain novel therapeutically useful compounds.

This highlight will examine the question if the recognition of a natural product by biosynthetic enzymes translates to recognition by the therapeutic target and whether such recognition would aid drug discovery.

2. Natural product approaches to drug discovery

The quest for new therapeutics has been approached by varying methods over time. Natural products represent one of the most enduring approaches to this goal. Historically, natural products from plants have been the basis for traditional medicine systems for thousands of years and according to the World Health Organisation (WHO) are still relied upon for primary health care by approximately 80% of the residents of developing countries in the world. Several reviews have been published which discuss this point. An example of the dominant role of natural products can be seen in the last 25 years where in the cancer therapeutic area, 77.8% of approved drugs are either natural products, based on natural products, or mimicks of natural products.

As of 2001 the number of globally used drugs derived from plants was estimated as 122 with 80% of these drugs having the same or related use as the plants from which they were derived. These figures indicate that only 20% of plant derived drugs were discovered independent of medicinal folkloric information. Today, in the face of increasing pressure to identify new lead compounds, several approaches to drug discovery from natural products are being explored. Screening techniques are widely employed today to discover new leads on the basis of the therapeutic target function. Recent times have seen the emergence of computing techniques to address, at the genomic scale, the target-ligand

recognition issue.

2.1. The screening approach benefits from combinatorial chemistry made by nature

The most common current approach to drug discovery from natural products is through screening. In most cases, collections of random samples are gathered and extracts tested for pharmacological properties. In the case of activity, the responsible natural product entity is identified and carried forward for further investigation. With the speed of High Throughput Screening (HTS) outpacing the rate of compound supply, the introduction of combinatorial chemistry some 20 years ago saw focus shift towards combinatorial chemistry with its promise of an abundance of compounds. Although the technology undoubtedly proved itself in its ability to efficiently provide large quantities of compounds for screening, combinatorial chemistry has not delivered the expected wealth of new drugs. Reports that only one *de novo* combinatorial compounds was approved as drug before 2006 [Newman and Cragg, 2007, J Nat Prod **70** 461_477] reflect this assessment. In the last decade, the question of quality over quantity with respect to combinatorial chemistry has thus become apparent.

However, combinatorial chemistry can, and has been employed to build upon the lessons learned by nature and using nature's own compounds as the basis for library development. There have been several comprehensive reviews of the types of combinatorial libraries which have been inspired by natural products. A review of libraries from natural product-like scaffolds in 2004 encompasses over fifty reported libraries based on carbohydrates, steroids, fatty acid derivatives, polyketides, peptides, terpenoids, flavonoids, alkaloids and other natural products. Aside from natural product libraries based upon combinatorial techniques there are other types of libraries based upon natural products which include pre-fractionated natural product libraries, pure natural product libraries, synthetic and semi-synthetic natural product libraries, focused libraries and natural product derived compound libraries based on diversity-modified natural scaffolds.

2.2. Chemical biology has brought new sources of inspiration

Recent times have seen more direct methods of harnessing the qualities of natural products based upon an understanding of the underlying specifics of their biological relevance. "Chemical biology" is a term that encapsulates target family concepts which focus on the ligand binding property of proteins to efficiently address the druggable genome. It is founded on the assumption that similar ligands bind to similar proteins. Two targets can be linked if they share common ligands, or alternatively, if they share common structural characteristics.

Associations among proteins on the basis of ligand chemical structure has allowed the prediction of new targets for known drugs. For instance, some side effects of emetin, a natural product extracted from ipecac root, could be explained by the identification of the α2-adrenegic receptor as an off-

target. While being a potent antiprotozoal agent, emetin may produce nausea, hypotension, tachycardia and congestive heart failure. The chemical structure of emetin revealed similarity to α2-adrenegic ligands, and the compound was indeed shown to modulate the α2-adrenegic receptor at 1 μM concentration, consistently with some of the adverse reaction of this drug.

A global analysis of the common ligands of a few hundred therapeutic targets revealed that many bioactive ligands are able to bind proteins assigned to different pharmacological classes. For instance, ligands of aminergic G-protein coupled receptors are often observed to target protein kinases; ligands of protein kinases have in turn been observed to possess unexpected activities on ion channels and phosphodiesterases. Since these targets are not homologuous proteins, they possess three-dimensional structural similarities in their ligand binding areas in the absence of significant sequence identity.

The direct comparison of protein three-dimensional structures has been successfully applied to target identification for natural product starting points. Early studies used the fact that a particular natural product binds to a protein with a particular fold, in order to develop compound libraries based on the natural product scaffold and targeted at other members of the same fold. The discovery of novel and highly selective inhibitors of tyrosine kinases vascular endothelial growth factor receptors 2 and 3 (VEGFR-2,-3), Tie-2 and insulin-like growth factor 1 receptor (IGF1R) was achieved by the selection of nakijiquinones, which are known inhibitors of the human epidermal growth factor receptor tyrosine kinase (Her-2/neu), as a scaffold for a screening library targeted at other receptor tyrosine kinases.

The ligand binding properties of a protein are however not systematically encoded in the overall fold, and a single fold may correspond to different protein functions, which themselves are associated to distinctive local structures. In an attempt to focus on the local structural similarity, the Waldmann group defined protein structure similarity clusters (PSSC). The ligand-sensing core is made of the secondary structure elements forming the ligand-binding domain, and thus consists in a subdomain of the overall fold. Compound libraries were developed to address a particular PSSC based on knowledge of natural product inhibitors of members of the cluster Proteins that share structural similarity despite low sequence identity are considered the most interesting cases. In a practical application this approach has afforded new inhibitors of members of the same PSSC; CDc25A phosphatase, acetylcholinesterase, and 11β-hydroxysteroid dehydrogenase type 1 and type 2. Inhibitors were contained within a 147-membered compound library based upon the natural product inhibitor of Cdc25A, dysidiolide, and the hypothesized importance of the γ-hydroxybutenolide group for the natural product's phosphatase inhibiting *activity the sentence has to be rewritten as the meaning is unclear* Though the enzymes screened in this PSSC are defined as having different folds according to SCOP (Structural Classification of Proteins Database), their active sites are nevertheless located in subdomains of common topology.

Recent considerations about the protein 3D similarity suggested that the conservation of structural fragments between proteins may be a signature of a common function, even in the absence of global sequence or fold similarity. The Honig group has thus deophanised a protein of unknown function from *Thermatoga maritima*. A scan of the Protein Databank identified structural neighbours defined by containing at least three aligned secondary structure elements with the query protein, thereby suggesting potential ligands. Boosted by the special interest of pharmaceutical companies to predict adverse drug reaction, the prediction of a ligand's target on the basis of protein local structure has become an active field of research. Because large scale studies demonstrated that two proteins can bind the same ligand even though they don't share sequence or fold similarity, methods and programs have been developed to address the protein local similarity issue independently of the sequence, the overall fold and the subdomain topology of proteins As a proof of concept, the Klebe group could explain the cross-reactivity of cyclooxygenase type-2 inhibitors with carbonic anhydrase. A common three-dimensional organisation of the pharmacophoric features exposed at the protein surface was observed in the two enzymes, although there was not a perfect match of shape and physicochemical properties of the enzyme active sites. Up to now, only two prospective studies successfully detected common binding sites among proteins unrelated in sequence and/or function. In both cases, a systematic search across the Protein Data Bank using the binding site of a given drug target as the query allowed the identification of a new off-target for the drug. In 2009, the Bourne group evidenced a relationship between the NAD-binding Rossmann fold of *M.tuberculosis* enoyl-acyl carrier protein reductase (InhA) and the SAM-binding domain of catechol-O-méthyltransferase (COMT), through similarities between their co-factor binding sites. Both theoretical protein-ligand docking and experimental antibacterial assays supported the computational predictions that synthetic COMT inhibitors, which are indicated for Parkinson's disease, also inhibit *M.tuberculosis* InhA. The next year, the Rognan group screened a library of 6,412 druggable protein-ligand binding sites thereby identifying an off-target for natural products used as inhibitor of protein kinases. Local structural similarity was observed between the ATP-binding sites of proto-oncogene Pim-1 kinase and synapsin I, a protein which regulates the neurotransmitter release in the synapse (Fig. 2). Binding and functional assays confirmed the predictions; The generic kinase inhibitor staurosporine was shown to bind synapsin I and to induce a functional response with submicromolar potency; The Pim-1 kinase selective inhibitor quercetagetin was as efficient yet twice more potent than staurosporine. No binding to synapsin I could be detected for a protein kinase A inhibitor or a protein kinase Chk1 inhibitor.

2.3 Detection of binding sites

In practice, binding site comparison algorithms involve three interrelated components; firstly the simplified representation of the site with various biochemical properties, secondly an algorithm to detect and superimpose common motifs, and

thirdly a method to score the resulting match. The quantification of local similarity between protein three-dimensional structures is generally thought of as a structural alignment problem; equivalences between two binding sites are established by optimising structure superposition. In brief, the methods used in this context so far include:

- ASSAM (Artymiuk *et al.*, 1994) reduces residue side chains to pseudo atoms and uses subgraph isomorphism to find common patterns between template and target.

- Cavbase (Schmitt *et al.*, 2002) and eF-site (Kinoshita & Nakamura, 2003) take a set of surface points and selected atoms (pseudo centres) of residue side chains to represent binding sites. Similarity is detected by subgraph isomorphism.

- SiteEngine (Shulman-Peleg *et al.*, 2004) also uses pseudo centres but performs the similarity search by geometric hashing which seems to be faster than the subgraph approach.

- SiteAlign (Schalon *et al.*, 2008) generates a cavity fingerprint on a 80 triangle-discretised sphere placed at the centre of the binding site and then performs a comparison of these fingerprints between template and target proteins.

- BSAlign (Aung & Tong, 2008) represents the binding site as graph including geomteric parameters as well as characteristics such as solvent accessibility, physicochemical type, secondary structure type, etc.

See the 2008 review by Kellenberger and colleagues for a detailed account of algorithms and available programs.

In the two examples described above (InhA/COMT and Pim-1 kinase/synapsin I), the protein representation was reduced to Cα atoms, which were defined with their 3D coordinates as well as the physico-chemical and geometrical properties of the corresponding residues.

Other methods describe the protein as the collection of the numerous pharmacophoric features at the binding site surface, thus including all hydrogen bond donors and acceptors, the positively and negatively charged groups, as well as the hydrophobic and aromatic groups). The pharmacophore-based approach has successfully identified targets for natural products. As an example, arborinine extracted from the medicinal plant *Ruta graveolens* was compared to 2208 structure-based and ligand-based pharmacophores, and found to match the pharmacophore derived from acetylcholinesterase (Fig. 2). [J. M. Rollinger, D. Schuster, B. Danzl, S. Schwaiger, P. Markt, M. Schmidtke, J. Gertsch, S. Raduner, G. Wolber, T. Langer, H. Stuppner, Planta Med. 2009, 75, 195–204.] The prediction was validated by *in vitro* binding assays.

/Esther to do the figure

Fig. 2 Chemical Biology approaches to drug design. The PCCS approach: the similar organisation of secondary elements in the ligand binding domain of Cdc25A phosphatase and 11β-hydroxysteroid dehydrogenase type 1 [Box 1]. The structural comparison of binding sites: the similar organisation of equivalent residues in the ATP-binding sites of Pim-1 and synapsin I [Box2]. The protein-based pharmacophore approach: . In the three boxes, the top picture shows the superimposition of the elements which represent the three-dimensional structures of query and target proteins; The bottom picture shows the ligand of the query protein demonstrated to be active on the target protein.

In principle, the ligand-binding preference of a query protein can be deduced from knowledge about the ligand binding preference to a target proteins provided a significant match is established between the protein-based pharmacophores of the query and the target proteins. Such an application has not yet been published.

4. Concepts around biosynthesis and biology space

The target-orientated approaches described involve using a specific natural product interaction with a target protein to guide identification of targets related by a specified degree of structural similarity. However, the natural product-target interaction is not the only interaction which can be exploited for target identification. The interactions made by natural products with their biosynthetic machinery presents another avenue for target identification.

4.1. Biosynthetic world

Natural products, or secondary metabolites, may be defined as naturally produced substances which do not play an explicit role in the internal economy of the organism that produces them. The reasons for their production is not entirely clear. It has been argued that organisms have evolved this ability to produce natural products because the functions of such compounds evidently serve the organism by improving its survival fitness. These roles include defending many species of plants, animals and micro-organisms against predators, pathogens and competitors

A particular natural product may serve its producing organism by acting as a toxin, repellent or attractant. The natural product exerts its effect through protein binding. From this we can conclude that nature has produced compounds which have the inbuilt capacity to interact with biology space. Despite the huge number of natural products being produced by organisms, for the vast majority, little is known about their actual role in nature. These compounds are manufactured in pathways consisting of an assembly line of biosynthetic enzymes, and thus have an imprint of their interaction with biology space from their very origin. Intriguingly, while natural products are not produced by nature to medicate humans, they do contain properties that make them favourable drug candidates. Natural products, having evolved in nature to interact with the biology space of proteins, have already accomplished what so many synthetic approaches aim to achieve. The reason that natural products are biologically validated starting points for drug design and hold the key to optimising the occupation of biology space is specifically due to their biosynthetic origin, the imprint of biology space and the common recognition elements within the limited protein structural arrangements.

4.2. Protein Folds

Within the protein world, there are a limited number of structural arrangements available and this knowledge has been leveraged. The majority of protein families belong to approximately 1000 common folds, describing the relative orientation of a particular set of secondary structure elements in three-dimensional space. The finite number of folds provides a simplification of biology space presented by

proteins. The question of interest is whether the interaction of natural products coded into their scaffold transfers into their interactions with a human target in a rational way.

4.3. The Protein Fold Topology (PFT) concept

As mentioned previously, examples exist of the application of the chemical biology approach to identify a new protein target for a natural product related by structural similarity. Our own efforts have been aimed at identifying what relationship is shared between a target and a biosynthetic enzyme of a natural product. Comparison made at the scale of local secondary structures revealed unexpected common organisation of the secondary structures forming the natural product binding sites. This structural similarity was defined as Protein Fold Topology (PFT).

The first PFT to be identified was the Flavonoid/kinase PFT. This investigation around flavonoids (fig. 3) was motivated by three main reasons: first the flavonoid biosynthetic pathway is well characterised, then flavonoids are known to inhibit a broad range of kinases, and last structures are available within the Protein Data Bank for both biosynthetic enzymes and therapeutic targets. The phosphoinositide 3-kinase was compared by eye to three biosynthetic enzymes (chalcone synthase, chalcone isomerase and anthocyanidin synthase) to identify the common secondary structure elements within a sphere of approximatively 15Å centered on the protein-bound ligand. The overlay of the PFT shared between anthocyanidin synthase and the phosphoinositide 3-kinase is illustrated in Fig. 4. Corresponding alignments and RMS deviations are given in Table 1. The quality of the RMS deviations for individual segments reflects the similarity in the positioning of secondary structure elements. By contrast with the ligand-sensing core of the PSSC approach, the PFT does not represent the fold of the active site domain, but consists in a limited number of secondary structure elements which form the active site and whose spatial arrangement is common to proteins which are unrelated by their sequence or their domain fold, but are related by the binding of common ligands, here flavonoids.

The PFT was also detected in the active sites of chalcone synthase and chalcone isomerase, with little variations in the number and the boundaries of considered secondary elements. Interestingly, the hydrogen binding pattern between quercetin and phosphoinositide 3-kinase is partly reproduced in the complex between quercetin and anthocyanidin synthase. As shown in Fig. 4, although the position of quercetin with respect to the PFT differs in the two enzymes, there is a good match between hydrogen bond interactions of the two enzymes with quercetin in select instances. The relative position of the key hydrogen bond donors and acceptors in chalcone synthase and chalcone isomerase, however, do not agree with those observed in phosphoinositide 3-kinase and anthocyanidin synthase. This is consistent with recently suggested multiplicity of flavonoid binding modes

The PFT study has demonstrated that the relationship between a natural product's biosynthetic enzyme and therapeutic target is not simple. The similarity is not obvious

and the subtleties of the relationship still need to be understood. In its simplest form, PFT uses the existence of secondary structure elements that make up the ligand binding site. Since ligands may interact with different main chain or side chain residues in the binding site when comparing template and target proteins, detailed atomic information as well as physicochemical properties are ignored in the first instance. Obviously, the position and orientation of the secondary structure elements has implications for the required shape complementarity (see Cavbase approach), but the PFT concept allows the substitution of one type of secondary structure element (helix, strand, etc) with another.

///
What is maybe missing here :
- Is the PFT a signature of flavonoid binding? In other words, it is a common motifs, also found in proteins which do not bind flavonoid at all ?
- Do other known flavonoid targets contain the identified PFT (eg. the estrogen receptor)?
People can rightly argue that geometric similarity may simply reflect structural and energetic constraints associated with the packing of secondary structure elements
//

In practice, the PFT approach lacks an automated identification tool, thus only a mature limited number of PFT relationships have been established to date. On this aspect, more research is necessary. A possible clue relies on the use of protein 3D-alignment methods. There are a number of algorithms available which perform different aspects of similarity searching; none of them, however, perfectly suits PFT identification.

Current computational methodologies can be classified into two categories. (i) Algorithms designed to perform three-dimensional superposition of a target and a template protein require reasonable structural homology globally or in a sequential section of the protein(s). While intrinsic restrictions caused by superposition of atomic coordinates by least squares methods are overcome by secondary structure matching [LIT] or alignment of gravitational axes [LIT], neither of the latter lends itself for the PFT problem. (ii) Algorithms exclusively comparing ligand binding sites address ligand-protein interactions at the level of atoms or residue side chains. The distinguishing feature of PFT, the occurrence of similar secondary structure positioning, is not considered.

PFT similarity does not require the same secondary structures to be compared, so that algorithms which effectively increase the structural similarity at the fold level of matched proteins will not include proteins that have similarity on the PFT level. Next, while proteins which share PFT may have similar binding sites, it is unknown if all proteins with similar binding sites will share similarity at the PFT level. For this reason, binding site comparison does not provide an answer for automated PFT searches.

Recent discussions of the evolution of proteins have suggested a continuity of the protein structural space, whereby
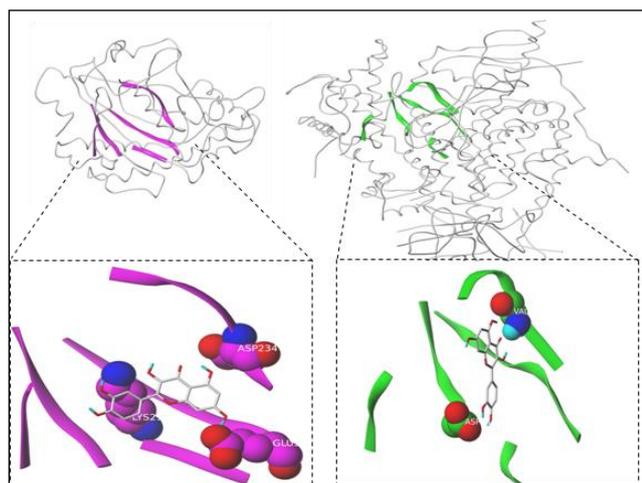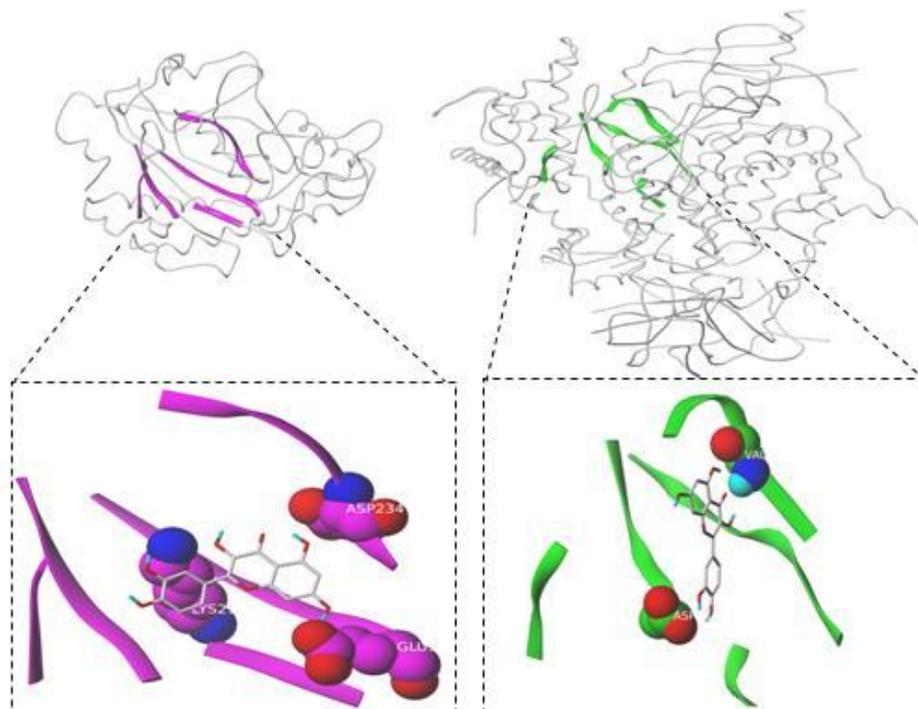
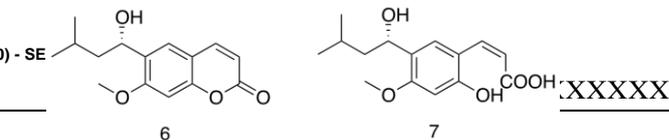**Table 1** Analysis of Flavonoid/kinase PFT alignment using LSQMAN

Aligned Residues		Number of Matched Residues	RMS for Residues (Å)
Therapeutic Target phosphoinositide 3-kinase (1e8w)	Biosynthetic Enzyme anthocyanidin Synthase (1gp6)		
Ala885-Glu880	Glu230-Val235	6	3.0
Leu960-Asp964	Glu140-Phe144	8	4.4
Asn949-Arg947	Ser127-Tyr125		
Cys863-Ile870	Ile299-Glu306	8	3.2
Arg839-Asp841	Leu210-Met212	3	5.7

4.4 Biosynthetic interactions reflected in target recognition

The biosynthesis of flavonoids involves a wide range of enzymes, which all exhibit distinctive characteristics in their active site. The core structure of flavonoid has therefore inherited a conformational flexibility as well as numerous combinations of hydrogen bond donors/acceptors to enable their interaction with all these biosynthetic enzymes (Here a figure could examplify the diversity of quercetin binding mode?), and also with other proteins with similar binding sites. The multiplicity of the flavonoid binding mode is consistent with the promiscuity of this natural product.

The flavonoid/kinase PFT study established a common protein architecture between some biosynthetic enzymes of flavonoid biosynthetic pathway and a protein of the kinase functional family. Binding site comparison across other biosynthetic enzymes indicates that the flavonoid recognition pattern

Fig. 6 The structure of 6-(1*S*-hydroxy-3-methyl- butyl)-7-methoxy-2H-chromen-2-one (6) and the ring open form (7) observed in the crystal structure with hCA II (5).

resulting in kinase recognition is not reflected in other biosynthetic enzymes. We expect that different biosynthetic enzymes will generate different interactions giving rise to additonal PFTs.

Therefore, to establish the universality of PFT, we need to identify other pairs of biosynthetic enzyme/therapeutic target with identical binding modes. This would establish that different biosynthetic enzymes can serve as starting point for drug design,.

In conclusion, we define the concept of "biosynthetic imprint" as the molecular recognition between substrate and biosynthetic enzyme and "natural product binding motif" as the molecular recognition between ligand and therapeutic target. A PFT

ESTHER _ PLEASE CONSIDER ABOVE

5. Implications for drug design

In the case of flavonoid metabolic enzymes/kinases, natural products of the flavonoid biosynthetic pathway represent starting points for drug design against protein kinases. The ability for flavonoids to function not only as kinase inhibitors, but as a potential kinase drug is demonstrated through the case of flavopiridol. Flavopiridol (Fig. 9) is the first potent inhibitor of cyclin dependent kinases and is currently in Phase III clinical trials } } fitoterapia 81 (2) 145-148 2010 While flavopiridol was synthetically derived from the plant alkaloid rohitukine isolated from the Indian plant *Dysoxylum binectariferum*. it contains the flavonoid scaffold.
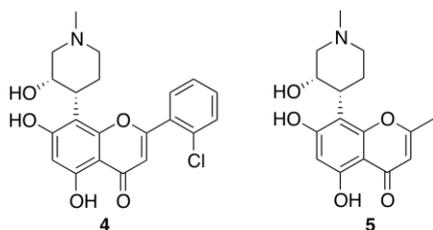

Fig. 5 The structure of flavopiridol (4) and rohitukine (5).

We have recently used bioaffinity mass spectrometry to screen natural product extracts for ligands that bind to bovine carbonic anhydrase II. We identified the coumarin (6) as a ligand for the enzyme. In subsequent work we determined that 6-(1*S*-hydroxy-3-methyl- butyl)-7-methoxy-2H-chromen-2-one (6) inhibited a range of human and murine carboninc anhydrases with greatest affinity against human CA II (K_I 59 nM) and CA I (K_I 80 nM). A crystal structure obtained by soaking 6 with human CA II showed that the ring open cinnamic acid 7 was present in the active site of human CA II.

As the biosynthetic enzymes of coumarin biosynthesis are known, we compared the binding mode observed in the human CA II ligand-bound crystal structure with docking of 6 and 7 into the biosynthetic enzymes. RESULT?

The identification of PFT provides another avenue for application of the chemical biology approach in drug discovery. Essentially, the automated comparison of PDB structures can serve three different aims (Fig. 10). Given a new therapeutic target, a similarity search could identify a biosynthetic enzyme, and the corresponding natural product would provide a starting point for drug design against the target. Similarly, a new natural product and its associated biosynthetic enzyme could be used to conduct a search for a potential target. Finally, PFT could be used with an established natural product-target relationship to direct a search for new targets. All three approaches require a substantial number of X-ray structures of biosynthetic enzymes.

a) natural product – identify biosynthetic enzymes – analyse for PFT – search PDB for same PFT – identify potential new target

b) Target – analyse for PFT – search for same PFT in biosynthetic enzyme – identify natural product as a potential lead

c) Natural product / target   - analyse for PFT – search PDB – identify new targets or potential off target effects

d) PFT to identify non-druggable targets

Fig. 10 The potential application of Protein Fold Topologies (PFTs) to discovering natural product starting points or new therapeutic targets. Up to 1,048 genes (3.5% of the genome) produce all proteins that are targets of current drugs. All current drugs with a known mode-of-action act through 324 distinct molecular drug targets. Of these, 266 are human-genome-derived proteins, and the remainder are bacterial, viral, fungal or other pathogenic organism targets. Small- molecule drugs modulate 248 proteins, of which 207 are targets encoded by the human genome. Oral small-molecule drugs target 227 molecular targets, of which 186 are human targets. In order to identify the familial relation- ships between all drug targets, the analysis domains, using the SCOP[19] and PFAM[20] databases showed that approximately 130 'privileged druggable domains' cover all current drug targets. This number is in stark contrast to the projected number of protein families and folds (10,000 folds[21] and more than 16,000 families[22]). One striking feature was that of the 361 new molecular entities (NMEs) approved by the FDA between 1989 and 2000, 76% targeted a precedented drugged domain and only 6% targeted a previously undrugged domain; the remainder have either unknown targets (4%) or are believed not to have distinct molecular targets underlying their action (17%). The rate of target innovation is surprisingly constant over the past 20 years, with an average rate of first-against-target drugs of 5.3 new 'drugged' targets per year.

A growing number of X-ray structures for biosynthetic enzymes

Since crystal structures of biosynthetic enzymes form the basis of the suggested PFT methodlgy, we appraised the their availability up to early 2006 is presented. An advanced search of the PDB was performed using the key words *biosynthesis* and *tailoring*. These particular key words were successful in returning previously identified biosynthetic enzyme crystal structures. A full text search scope was selected and the search returned 2566 structures for the first keyword and 6 for the second. The search also retrieved some entries such as enzymes involved in "protein biosynthesis" which were deemed irrelevant and not included unless further clarification using KEGG warranted their inclusion. The results were refined by analysis of each of the PDB entries on the basis of the enzyme's biological role.   Biological roles were determined from information in the PDB stored under *biological process* or as key words under *biology and chemistry*. Abstracts and full text articles together with the KEGG (Kyoto Encyclopedia of Genes and Genomes) PATHWAY database were referred to for clarification where necessary and for grouping enzymes into biosynthetic pathways.

 The search of the Protein Data Bank revealed 2385 crystal structures of biosynthetic enzymes which consist of 351 known different enzymes in the biosynthetic pathways of 158 different natural products. A summary of the results of the search are tabulated in Table 6 (see Supporting Information for the complete search results including individual PDB IDs). Table 6 shows the different pathways for which biosynthetic enzymes have been solved together with the total number of PDB entries and where known, the number of unique contributing EC numbers (Enzyme Commission numbers) which indicates different biosynthetic enzymes.

Table 6 Summary of biosynthetic enzyme X-ray structures in the PDB

Pathway (KEGG)	Unique EC numbers (~)	Number of PDBs
Amino Acid Metabolism	111	584
Biosynthesis of Polyketides and Nonribosomal Peptides	5	44
Biosynthesis of Secondary Metabolites	24	107
Carbohydrate Metabolism	9	51
Energy Metabolism	1	3
Glycan Biosynthesis and Metabolism	32	122
Lipid Metabolism	50	334
Metabolism of Cofactors and Vitamins	68	430
Metabolism of Other Amino Acids	2	15
Nucleotide Metabolism	42	499
Xenobiotics Biodegradation and Metabolism	2	11
Undeterminable	11	98

Notably, 24 X-ray structures for different biosynthetic enzymes of various secondary metabolites have been solved. 5 This first sentence is unclear in its meaning. The

Biosynthesis of Secondary Metabolites is broken down in KEGG into 16 further classifications; Please don't list in text - it cannot be digested. Can this be included in the table above or over the 10 page? Terpenoid biosynthesis; Diterpenoid biosynthesis; Monoterpenoid biosynthesis; Limonene and pinene degradation; Indole and ipecac alkaloid biosynthesis; Stilbene, coumarine and lignin biosynthesis; Flavonoid biosynthesis; Alkaloid biosynthesis I; Alkaloid biosynthesis II; Penicillins 15 and cephalosporins biosynthesis; beta-Lactam resistance; Streptomycin biosynthesis; Tetracycline biosynthesis; Clavulanic acid biosynthesis; Puromycin biosynthesis; Novobiocin biosynthesis. The main pathways contained within the KEGG pathway *Biosynthesis of Secondary* 20 *Metabolites* are shown in Fig. 11. The X-ray structures available within this pathway and the natural products biosynthesised are given in Table 7.

Table 7 X-ray structures for biosynthetic enzymes of secondary metabolites

Pathway	PDB ID	ecNo	Biosynthetic Enzyme Descriptor	Natural Product
Alkaloid biosynthesis II	1AE1	1.1.1.236	tropinone reductase-I	tropine
Alkaloid biosynthesis II	1IPE, 1IPF, 2AE1, 2AE2	1.1.1.236	tropinone reductase-II	tropine
Clavulanic acid biosynthesis	1UPA,1UPB, 1UPC	2.5.1.66	carboxyethylarginine synthase	L-N2-(2-carboxyethyl)arginine
Clavulanic acid biosynthesis	1GQ6, 1GQ7	3.5.3.22	proclavaminate amidino hydrolase	proclavaminate
Erythromycin biosynthesis	1Z8O, 1Z8P, 1Z8Q	1.-.-.-	wild-type cytochrome P450eryF	6-deoxy erythronolide B
Erythromycin biosynthesis	1PZQ, 1KEZ, 1PZR	2.3.1.94	erythromycin polyketide synthase (DEBS)	6-deoxy erythronolide B
Erythromycin biosynthesis	1OXA	n/a	cytochrome P450	erythronolide B
Flavonoids, stilbene and lignin biosynthesis	1YQD, 1YQX	1.1.1.195	sinapyl alcohol dehydrogenase c	sinapoyl aldehyde/alcohol, coniferyl aldehyde/alcohol, 4-hydroxy-cannamyl aldehyde/alcohol
Flavonoids, stilbene and lignin biosynthesis	1GP4, 1GP5, 1GP6	1.14.11.19	anthocyanidin synthase	dihydroquercetin
Flavonoids, stilbene and lignin biosynthesis	1SUI, 1SUS	2.1.1.104	alfalfa caffeoyl coenzyme A 3-O-methyltransferase	feruloyl-CoA
Flavonoids, stilbene and lignin biosynthesis	1KYW, 1KYZ	2.1.1.68	caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase in complex with 5-hydroxyconiferaldehyde	sinapate, ferulate
Flavonoids, stilbene and lignin biosynthesis	1BI5, 1BQ6, 1CGK, 1CGZ, 1CHW, 1CML, 1D6F, 1D6H, 1D6I, 1EE0 1I86, 1I88, 1I89, 1I8B, 1JWX, 1U0V, 1U0V, 1U0W	2.3.1.74	chalcone synthase	naringenin, isoliquiritoigenin
Flavonoids, stilbene and lignin biosynthesis	1QLV	2.3.1.74	pyrone synthase	naringenin, isoliquiritoigenin
Flavonoids, stilbene and lignin biosynthesis	1EYP, 1EYQ, 1FM7, 1FM8, 1JEP, 1JX0, 1JX1	5.5.1.6	chalcone isomerase	naringenin, isoliquiritoigenin
Flavonoids, stilbene and lignin biosynthesis	1ZGD	n/a	chalcone reductase	deoxychalcone
Novobiocin biosynthesis	2C0Z	n/a	novW: A 4-keto-6-deoxy sugar epimerase	novobiocin
Penicillins and cephalosporins biosynthesis	1W2N, 1DCS, 1E5H, 1E5I, 1HJF, 1HJG, 1RXF, 1RXG, 1UO9, 1UOG	1.14.20.1	deacetoxycephalosporine C synthase	deacetoxy-cephalosporin C
Penicillins and cephalosporins biosynthesis	1ODM, 1UZW, 1W03, 1W05, 1W06, 1W3V, 1W3X, 2BU9, 1BKD, 1ODN, 1QIQ, 1QJE, 1QJF, 1BLZ, 1HB1, 1HB2, 1HB3, 1HB4, 1IPS, 1OBN, 1OC1	1.21.3.1	isopenicillin N synthase	isopenicillin N
Streptomycin biosynthesis	1BWD	2.1.4.2	inosamine-phosphate amidinotransferase STRB1	1-guandino-1-deoxy-scyllo-inositol-4P
Streptomycin biosynthesis	1GR0, 1LA2	5.5.1.4	myo-inositol phosphate synthase.	myo-inositol-1P
Terpenoid biosynthesis	1UBV, 1UBW, 1UBX, 1UBY	2.5.1.1	farnesyl pyrophosphate synthetase	geranyl-PP, farnesyl-PP
Terpenoid biosynthesis	1FPS, 1RQI, 1RQJ, 1YQ7, 1YV5	2.5.1.10	farnesyl diphoshate synthase	geranyl-PP, farnesyl-PP
Terpenoid biosynthesis	1EZF	2.5.1.21	squalene synthase	squalene
Terpenoid biosynthesis	1HZT, 1I9A, 1P0K, 1P0N, 1PPV, 1PVF, 1R67, 1X83, 1X84	5.3.3.2	isopentenyl diphosphate isomerase	dimethylallyl-PP, isopentenyl-PP

While the number of elucidated structures of biosynthetic enzymes is growing with increasing speed, there is still enormous potential for structure determination especially for enzymes not involved in major pathways and for those which play a role in the final stages of biosynthesis for interesting natural products. Given the correlation found between the target and biosynthetic enzymes of natural products, value should be placed on structure determination of biosynthetic enzymes to allow for further exploration of this new avenue to drug design.

Conclusions

Synthesis in the biosphere, *per se*, does not endow a natural product with the property of modulating the function of proteins of therapetic potential. The fact that natural products have already experienced an interaction with a protein also does not, *a priori*, impart any ability to sense a protein surfaces at sites that modulate therapeutic targets.

In the case where the intended target protein of a natural product has a human homologue which is involved in disease, it is clear that the natural product can affect the human target. However, this does not explain the occurrence of high numbers of natural products that interact with therapeutic targets.

Using the new methodology of PFT to describe the imprint of the biological interactions during biosynthesis it is suggested that this imprint does indeed correspond to interactions with therapeutic targets.

An important reason may be the conservation of

We conclude from PFT that, as all secondary metabolites are engineered by interacting with enzymes through biosynthetic pathways, all such compounds represent potential starting points for drug design.

(ESTHER ?)

We have reviewed the possibility to use the biological imprint during biosynthesis (PFT) as a drug discovery tool with respect of identifying molecular targets rather than relying on screening methods.

Natural products provide a chemical diversity set that contains compounds that are complementary to biological structure space (Fig. 1).
The Protein Fold Topology correlations discovered between a natural product's biosynthetic enzyme and therapeutic target explains why so many natural products have become therapeutic agents. PFT defines a natural product's ability to recognise biology space and represents a potential tool for drug design.

Expand conclusion section - reiterate main points and utility of PFT and compare to alternative approach of PSSC (and others). Is your method limited to the druggable genome or is there potential to use it to identify metabolites that might be used to interfere with protein-protein interactions?
What are the benefits/advantages and limitations of your method.

Acknowledgements

Notes and references

a Eskitis Institute, Griffith University, Nathan, Brisbane, QLD 4111, Australia. Fax: +61 07 3735 6001; Tel: +61 07 3735 6006; E-mail: r.quinn@griffith.edu.au
b Université de Strasbourg, F-67400 Illkirch, France