The clustering of proteins based on the structural similarity of their ligand-sensing cores combined with natural product-guided compound library synthesis provides a new rationale for chemogenomics.

Protein structure similarity clustering and natural product structure as guiding principles in drug discovery

Marcus A. Koch and Herbert Waldmann

The identification of new chemical entities that are capable of altering protein function lies at the heart of the hit and lead finding process, for which combinatorial chemistry has emerged as a powerful tool. Following the maturation of combinatorial chemistry and compound library development, it was soon recognized that biological relevance, design and diversity of a library are more important than library size. The universe of conceivable compounds is almost infinite, therefore, the decisive question arises: where is a biologically validated starting point in structural space from which to build a compound library to be found? As a new approach to address this complex problem, a synergistic strategy is presented, which is based on protein structure similarity clustering and natural product structure as guiding rationales.

During the past few decades, information about biological systems has grown rapidly, in particular through large-scale and global approaches addressing DNA sequence (genomics), protein structure (structural genomics) and protein expression and interactions (proteomics). These developments raise the expectation that the initial set of basic data will be converted to knowledge resulting in the development of novel therapies and drugs. The rapidly growing knowledge related to possible new drug targets has been paralleled by the development of combinatorial chemistry techniques that enable rapid synthesis of large compound libraries to be investigated in drug discovery programs.

Starting points for compound library development

The ostensibly high productivity, in terms of compound numbers, using the most advanced technology accessible today, has not yet led to a corresponding increase in the number of new drug candidates.

Marcus A. Koch
Herbert Waldmann*
Max Planck Institute of Molecular Physiology, Department of Chemical Biology, and Fachbereich 3, Organische Chemie, University of Dortmund, D-44227 Dortmund, Germany
*e-mail: herbert.waldmann@mpi-dortmund.mpg.de

MARCUS A. KOCH
Marcus A. Koch, born in 1975, studied pharmacy at the University of Freiburg as scholar of the Studienstiftung des deutschen Volkes. Following pharmacy training at City-Apotheke, Freiburg, he completed his pharmaceutical education, achieving the state license as a pharmacist in 2002, and joined the department of Chemical Biology, Max Planck Institute of Molecular Physiology, Dortmund, Germany. He is currently pursuing doctoral studies under the guidance of Herbert Waldmann as scholar of the Studienstiftung des deutschen Volkes. He is interested in interdisciplinary approaches in drug discovery, integrating bio- and chemo-informatics, organic synthesis and biochemistry.

HERBERT WALDMANN
Herbert Waldmann, born in 1957, received his Dr. rer. nat. degree in 1985 (University of Mainz, Germany). After postdoctoral (1985–1986) studies at Harvard University, under the supervision of George Whitesides, and habilitation (1991; University of Mainz), he accepted a professorship at the University of Bonn in 1991.
In 1993, Waldmann moved to the University of Karlsruhe as Full Professor of Organic Chemistry. In 1999, he was appointed Director at the Max Planck Institute of Molecular Physiology (Department of Chemical Biology), Dortmund, and Full Professor of Biochemistry at the University of Dortmund.
Herbert Waldmann has been the recipient of the Friedrich Weygand Award for the advancement of peptide chemistry, the Carl Duisberg Award of the Gesellschaft Deutscher Chemiker, the Otto-Bayer-Award, the Steinhofer Award of the Steinhofer Foundation and the Max Bergmann Medal. His current research interests include chemical biology, natural product synthesis, stereoselective synthesis, combinatorial chemistry and the chemistry and biology of lipid-modified proteins.

1359-6446/04/$ – see front matter PII: S1359-6446(04)03419-7

Instead, after more than a decade of combinatorial chemistry research, it has become evident that increasing the size of compound collections alone is not sufficient to increase the productivity of drug development. Rather, one of the most important questions in modern drug development still remains basically unanswered: where in chemical structural space are biologically relevant compounds to be found?

The answer to that question is certainly not trivial. However, some criteria have already been identified as being central in designing compound libraries for protein ligand development: 'diversity' [1–4]; 'drug-likeness' [5–9]; and 'biological relevance' [10–12]. This last criterion might be fulfilled if so-called 'privileged' [13] structures or natural products that have evolved to bind to biological macromolecules are taken as guiding structural principles. Indeed, hit rates increase significantly if natural compounds, or analogs thereof, are included in HTS [10,11,14,15]. Considering potential target proteins, several concepts – predominantly based on a clustering of target proteins according to evolutionary relatedness and conserved molecular recognition – have been developed and applied to steer ligand development. The major limitation of these approaches is that they usually consider close sequence homologs. However, for proteins, spatial structure is more conserved in evolution than amino acid sequence [16]. This suggests focusing more on the principal architecture and structure of proteins when conceiving alternative guiding principles for the development of biologically relevant compound collections.

Protein domain architecture

The structural conservatism of nature in the design of proteins that can be regarded as 'modularly built biomolecules assembled from individual domains as building blocks' could provide opportunities to develop guidelines for the identification of biologically relevant ligand structures. Domains are defined as discrete parts of a protein that fold independently from the rest of the structure to a compact arrangement of secondary structures interconnected via more or less complex linker peptides. Different sequence families (domains) can adopt the same fold, either as a result of convergence caused by functional and physical constraints because of the limited number of acceptable spatial arrangements of secondary structural elements, or as a result of divergent evolution to an extent that the sequence relationship is no longer recognizable [17,18]. Protein domains can be regarded as structurally conserved yet genetically mobile units [19]. Thus, the core structures of protein domains – the catalytic or the ligand-sensing cores – are widely reused in different functional contexts in a more or less modified form. Often, the binding- or catalytic-sites in proteins are diverse, whereas the domain cores are structurally conserved. Although the estimate for the number of different proteins in humans ranges from 100,000 to 450,000, there is a common agreement that the number of domain families, and furthermore of topologically distinct folds, will be much smaller. At present, according to the structural classification of proteins (SCOP) database [20], ~800 folds are known [21], which are derived by classifying all structurally characterized proteins according to their 3D structure and evolutionary arguments. Data from the ongoing genome sequencing projects enable an estimation of the number of existing folds and protein families in nature. Current estimates vary from 650 to 10,000 distinct folds, and 4000–50,000 sequence families [18,22–28], and it seems certain that the majority of protein families belongs to ~1000 common folds [29].

In this context, it should be mentioned that, in many cases, fold definition remains an empirical approximation and even experts disagree on fold assignments for many proteins. This is mainly because the criteria used are often rather loose. Frequently, not only structural data but also evolutionary and functional considerations are taken into account. Instead, categorization along exclusively structural aspects would be more appropriate because proteins of the same fold do not necessarily share a common ancestor or have similar physiological roles.

There is an ongoing effort to reveal the correlation patterns of protein function and protein sequence [30–32]. Although we are still far away from a deep and consistent understanding, and analysis is hampered by the small number of available X-ray structures of proteins with bound small-molecule ligands, some interesting observations that are relevant to the theme discussed here have been made about the diversity and evolutionary relationships of ligand-binding sites in proteins [33–35].

The ligand-sensing cores of protein domains

The ligand-binding or catalytic sites are the most relevant subsets of a protein domain from the point of view of development of small-molecule binders. These are, more precisely, located within the so-called ligand-sensing core of the domain where the actual catalytic conversion of enzymes, or the binding event of small-molecule ligands, occurs. The definition of these topologically distinct domain parts is all the more important when ligand-binding domains are quite large compared with the structural subset relevant for binding. Thus, we suggest confining structural similarity considerations to these distinct parts of a protein domain and grouping ligand-sensing cores, instead of whole domains, according to 3D similarities into so-called protein structure similarity clusters (PSSCs). This further broadens the structural view on proteins because it also enables consideration of structural parts of a previously defined fold. Thus, some structural fold characteristics could be reused within a different fold. When these similar fold parts describe the ligand-sensing cores of a domain, then a comprehensive clustering according to purely structural arguments might be a viable abstracting rationale in the context of ligand development, irrespective of more or less arbitrarily assigned fold types.

Protein structure similarity and natural product frameworks as guiding principles for compound library development

Classically, potential protein targets are clustered into target families on the basis of functional relatedness and amino-acid-sequence homology alignments reflecting their evolutionary relationship. This categorization is then used to pool known ligands of a target family and to take them as starting points for compound library design. This strategy constitutes a rationale that enables the direct conversion of genetic information and relatedness into actual chemical ligand design. A further, analogous principle was outlined as the 'structure–activity relationship homology concept'. Potential drug discovery targets are grouped into families based on the relatedness of the SAR of their ligands [36]. It is assumed that the conservation of binding-site architectures and, thus, the relatedness of molecular recognition within a target family or a subfamily thereof translate into a conservation of ligand scaffolds that bind to these targets [37]. The major limitation of these concepts is that usually only close sequence homologs can be considered because target proteins and their ligands are predominantly categorized on the basis of function and sequence similarity. Family assignment derived from sequence information alone in the absence of structural information usually requires sequence identities >30% [34].

Protein spatial structures are typically more conserved in evolution than amino acid sequences [16]; therefore, these should be considered first and foremost for the clustering of target proteins. Usually, structural conservatism in nature is confined to the domain cores of proteins, whereas the binding sites for ligands might be structurally diverse yet topologically similarly located. This concurrence of binding sites rarely indicates any obvious functional similarity because substrates can show high variability concerning their chemical structure. This observation could be attributed to a general tendency of particular folds to bind substrates at similar locations, at so-called 'supersites', despite little evidence of a common ancestor for the proteins considered. Possible explanations for this phenomenon might be particular principles of protein structure or chemical constraints that could lead to common 'optimal' binding sites, even when proteins do not share a common ancestor [38].

These observations indicate that a clustering of ligand-sensing or catalytic cores of proteins exclusively based on structural considerations could be a valuable tool for compound library development, thereby taking into consideration the structural conservatism in nature. Thus, we propose that protein domain cores be grouped in a PSSC because of 3D similarity and regardless of significant sequence similarity. A substance that binds to one member protein of a PSSC could be used as a starting point for the development of small-molecule modulators of the other members of the cluster. The biological diversity occurring in the ligand-binding sites has to be addressed with an appropriate chemical diversity that must be generated around the identified structural framework in a library approach to find ligands with an acceptable frequency in the first instance, and also to achieve potency and selectivity. Thus, an initial focus on the conserved architecture of protein domains or domain cores is used as an abstracting guiding principle, which leads to biologically relevant frameworks to be used as cornerstones of a compound library with significantly enhanced hit rates [11,12,39] (Figure 1).

The PSSC approach, synergistically applied with natural product-inspired combinatorial chemistry, is presented here as a strategy for the development of compound libraries for chemical biology and medicinal chemistry research. Natural products can be regarded as small molecules that have been evolutionarily selected for binding to protein domains. They interact with multiple proteins in the course of their biosynthesis and they target further proteins when they fulfill their biological functions, for example, in communication or chemical defense. This is further supported by the finding that major classes of natural products show multiple biological activities. Therefore, because of their biological pre-validation and evolutionary proving, they are particularly well-suited as structural 'leitmotifs' for compound library design. However, not only natural products fulfill the criterion of biological pre-validation to serve as promising guiding structures for library design. Various non-natural product classes, discovered, in particular, in medicinal chemistry programs, also have this property.

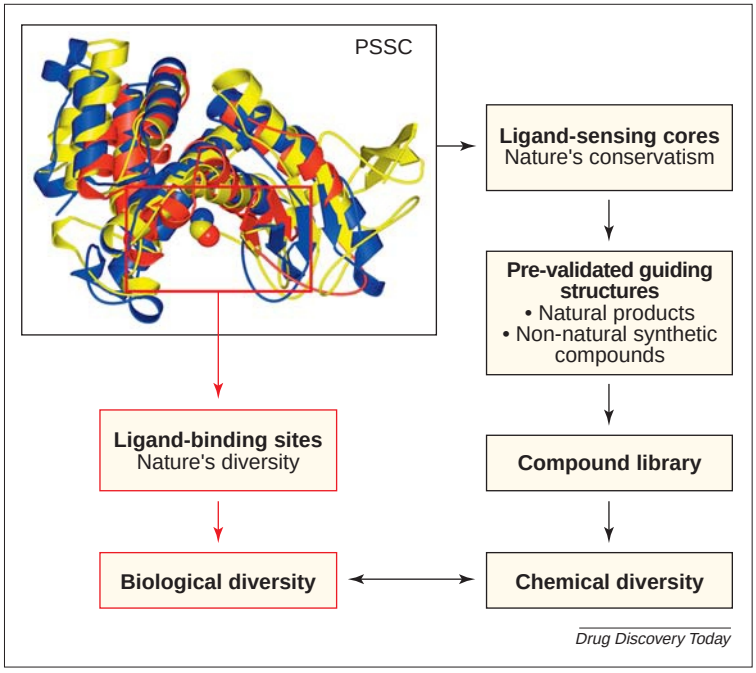

FIGURE 1

Protein fold conservatism and binding site diversity: implications for compound library development. Structural conservatism in the ligand-sensing cores of proteins leads to the identification of guiding structures for ligand development. However, to address the biological diversity in the ligand-binding sites, an appropriate chemical diversity has to be generated around the identified guiding scaffolds in a library approach.

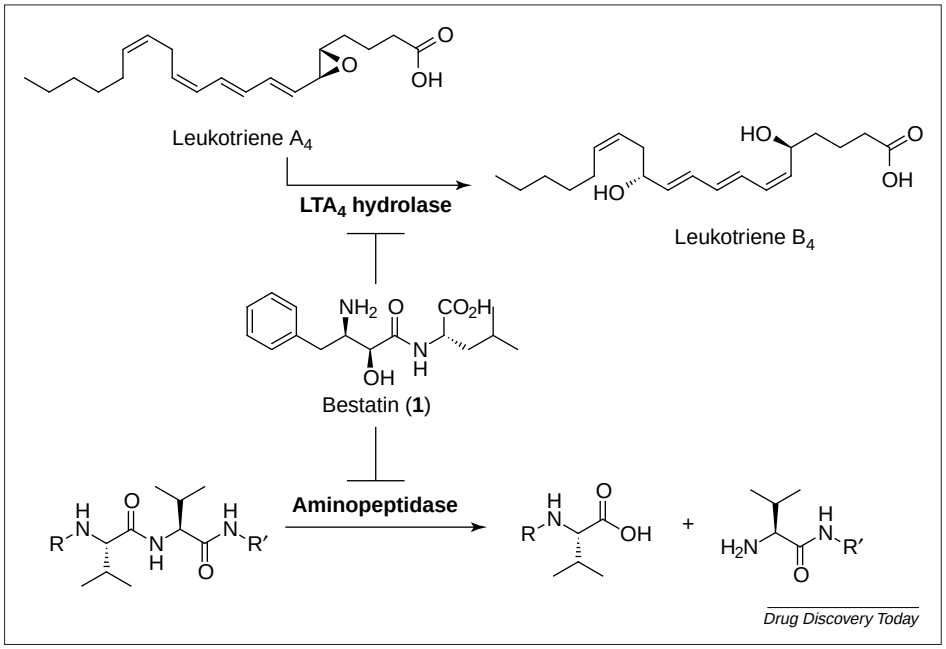

FIGURE 2

Reactions catalyzed by leukotriene A_4 hydrolase and aminopeptidases. Although LTA_4H and aminopeptidases are inhibited by the natural product bestatin, a known aminopeptidase inhibitor, they catalyze two different reactions. LTA_4H carries out the vinylogous hydrolysis of the leukotriene epoxide LTA_4 into LTB_4 in its zinc-containing active site, whereas aminopeptidases cleave peptide bonds.

PSSC-guided small-molecule binder development, in principle (although not necessarily desirable), could be initiated exclusively using bioinformatics tools without further knowledge about the biological functions of the target, binding partners and so on, which are usually obtained by laborious biochemical and cell biology techniques. Indeed, the evolved ligands can be used for further characterization of the physiological role of the target protein, which is of considerable importance in the target validation process and in chemical biology research.

Here, we provide evidence for the applicability of the PSSC concept in compound library development. First, known investigations extracted from the literature that led to the discovery of small-molecule modulators of protein function using other strategies are presented and re-analyzed in light of the PSSC concept. Second, the first successful application of this rigorously abstracting PSSC approach

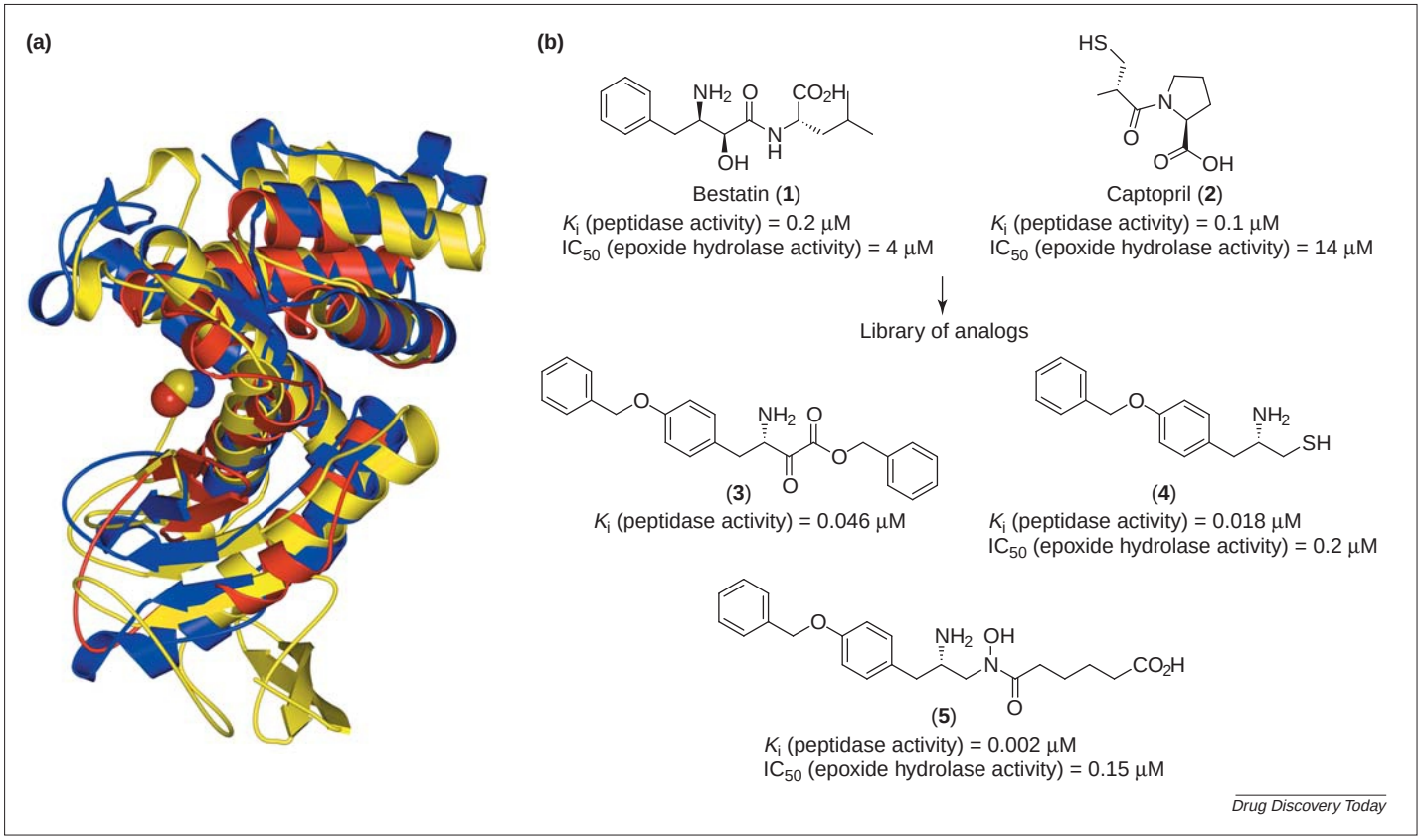

FIGURE 3

Superimposition of the X-ray structures of the catalytic domains of leukotriene A_4 hydrolase, angiotensin-converting enzyme and thermolysin. (a) The catalytic cores of LTA_4H (blue), ACE (red) and thermolysin (yellow), including the catalytic zinc ions (colored accordingly), superimpose well. The presence of the zinc-binding motif (HEXXH-X_{18}-E) in LTA_4H was sufficient to prompt investigations of the relationship of this enzyme to zinc-binding metallopeptidases. LTA_4H is assigned to the superfamily of metalloproteases according to SCOP [20]. **(b)** Bestatin- and captopril-derived inhibitors of LTA_4H.

as a rationale for actual *de novo* compound library design is discussed.

PSSC of leukotriene A_4 hydrolase and aminopeptidases: development of leukotriene A_4 hydrolase inhibitors

Leukotriene A_4 hydrolase (LTA_4H) is a bifunctional zinc metalloenzyme that converts LTA_4 into LTB_4 and also exhibits aminopeptidase activity. LTB_4 is a potent chemoattractant and immune-modulating lipid mediator involved in inflammation, immune responses, host defense against infection and platelet-activating factor (PAF)-induced shock. The crucial role of LTA_4H in LTB_4 generation makes it an attractive drug target. LTA_4H performs the vinylogous hydrolysis of the leukotriene epoxide LTA_4 into LTB_4 in its zinc-containing active site [40,41]. The zinc ion serves as a Lewis acid, polarizes the epoxide ring and stabilizes the negative charge occurring in the transition state. In the case of LTA_4H, the presence of the zinc-binding motif (HEXXH-X_{18}-E) was sufficient to prompt investigations of the relationship of this enzyme to zinc-binding metallopeptidases [42]. The evolutionary relationship of the LTA_4H fold to metallopeptidases would have immediately suggested searching for peptidase inhibitors as potential ligands, and indeed the naturally occurring aminopeptidase inhibitor bestatin (**1**; Figure 2) also inhibits LTB_4 biosynthesis. This result, and the related observation that the angiotensin-converting enzyme (ACE) inhibitor captopril (**2**; Figure 3b) also inhibits LTA_4H [42], have inspired the combinatorial variation of these lead structures, which led to the syntheses of potent inhibitors of the epoxide hydrolase activity of LTA_4H that also exhibited selectivity for LTA_4H, when compared with the inhibitory effect towards other aminopeptidases (compounds **4** and **5**; Figure 3b) [43–47].

The catalytic cores of LTA_4H (classified as a member of the M1 metallopeptidase family), human ACE (a member of the M2 family) and thermolysin (which belongs to the M4 family) share little sequence similarity (sequence identities amount to ~7%). However, they do exhibit significant structural resemblance. Thus, retrospectively, the catalytic cores of these three enzymes would have been grouped into a PSSC (Figure 3a) based on purely structure-driven considerations. The catalytic sites are similarly located, as indicated by the overlapping position of the catalytic zinc ions of all three enzymes in space. Interestingly, in this ligand design project, a natural product, bestatin (**1**), a non-specific aminopeptidase inhibitor, constituted the initial structural rationale for the development of LTA_4H inhibitors. This lends proof to the idea that natural products are particularly well-suited as biologically validated starting points in structural space for the design of protein ligands.

PSSC of nucleotide kinases and PAPS sulfotransferases: development of sulfotransferase inhibitors

To target cyclin-dependent kinases (CDKs), a purine scaffold-based compound library inspired by the natural product olomoucine (**6**; Figure 4b) was developed by Norman *et al.* [48] and Gray *et al.* [49]. This library afforded a moderately potent inhibitor of CDK2. Further development of this library by synthesizing several hundred 2,6,9-trisubstituted purine derivatives using solid- and solution-phase chemistry yielded more potent CDK inhibitors (CDK1 and CDK2), such as compound **7** (Figure 4b) [49,50]. A screen of this representative library of purines against recombinant inositol-1,4,5-trisphosphate-3-kinase (IP3K) led to the discovery of inhibitors of IP3K (e.g. **8**; Figure 4b) [51].

Kinases and sulfotransferases catalyze transfer reactions of anionic groups and use structurally similar cofactors. Both enzyme classes are capable of binding adenosine-based cosubstrates. Kinases bind adenosine-5′-triphosphate (ATP) as a phosphoryl donor and sulfotransferases bind 3′-phosphoadenosine-5′-phosphosulfate (PAPS) as a sulfate donor.

Interestingly, the catalytic cores of PAPS sulfotransferases and nucleotide kinases, as shown in the superimposition of yeast uridylate kinase (yUK) with murine estrogen sulfotransferase (mEST) (Figure 4a), display significant structural similarity, although the catalytic domains exhibit insignificant sequence similarity (sequence identities amount to 8%). Both proteins bind their cofactors through backbone amide hydrogen-bond interaction using a P-loop motif to bind the penultimate phosphate, whereas the specific side chain interactions differ. In addition, the phosphate on the substrate that is phosphorylated by yUK has the same orientation with respect to the cofactor as the phenolic hydroxy group of 17β-estradiol that is sulfated in the mEST-catalyzed reaction. This suggests that the catalytic mechanism of sulfuryl and phosphoryl transfers might be similar. Despite the high degree of structural similarity in the catalytic cores, there are only a few conserved amino acids, and the specific residues involved in catalysis derive from different locations in the active site [52,53].

The similarities concerning the bound cofactors, the reaction mechanism and the adenine-binding pockets led to a screen of the previously described purine-based library of ATP-competitive inhibitors, which were originally designed to target CDKs for cross-reactivity with the carbohydrate sulfotransferase NodH from *Rhizobium meliloti*. PAPS-competitive NodH inhibitors (e.g. **9**; Figure 4b) with modest inhibitory activity (IC_{50} values of 20–40 μM) could be identified that showed selectivity among several tested sulfotransferases, whereas all sulfotransferase inhibitors also displayed inhibitory activity in the μM range against several kinases [54].

This compound collection was also tested with mEST. This screen afforded a purine-based inhibitor of mEST with nanomolar potency that showed weak activity against several CDKs, yet selectivity for mEST when tested with representative members of the carbohydrate sulfotransferase family [55]. Finally, a screen of the purine library against β-arylsulfotransferase-IV (β-AST-IV) led to the discovery of a potent and highly selective inhibitor (**10**; Figure 4b) of β-AST-IV (K_i of 96 nM). This compound was

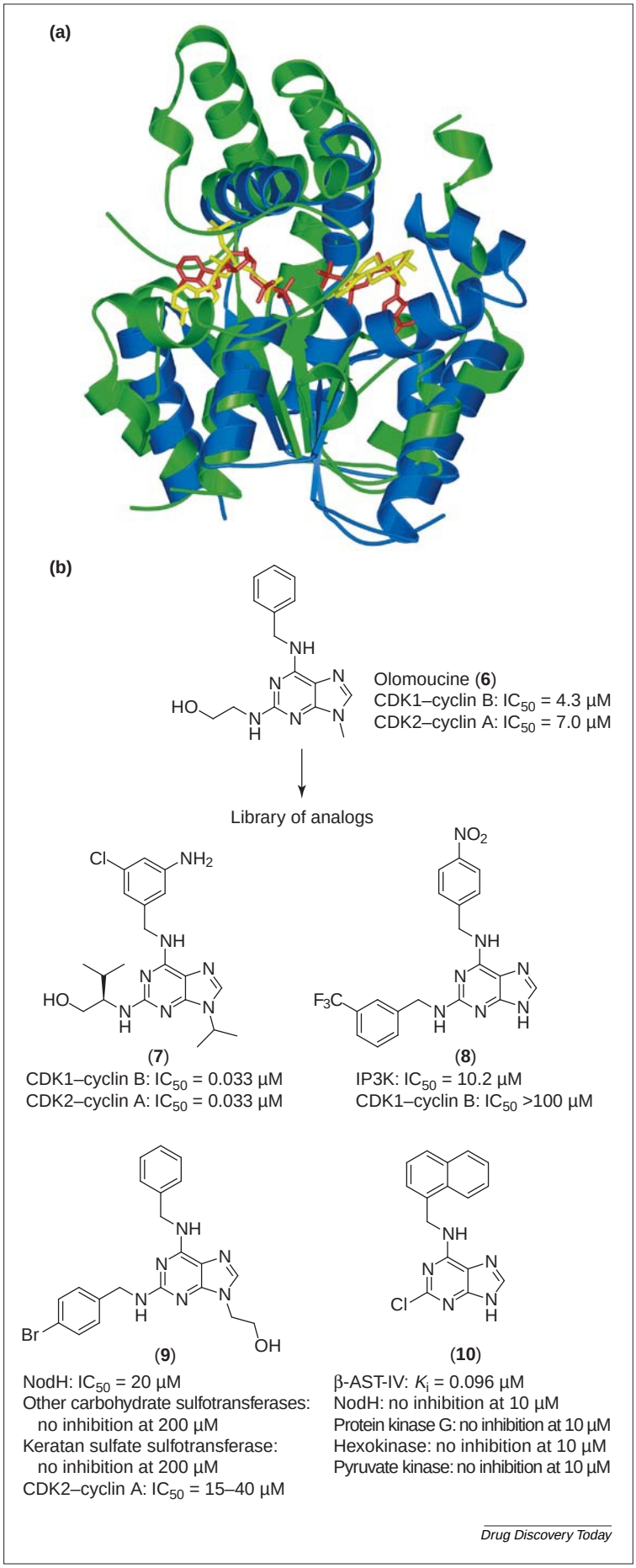

FIGURE 4

Superimposition of the X-ray structures of the catalytic domains of estrogen sulfotransferase and yeast uridylate kinase in complex with their cofactors and substrates. (a) EST (green) with consumed cofactor (PAP) and substrate (17β-estradiol) in yellow; UK (blue) with consumed cofactor (ADP) and substrate analog (ADP) in red. **(b)** Representatives of olomoucine-based libraries yielding inhibitors that target kinases and PAPS sulfotransferases.

also profiled against a variety of nucleotide binding proteins (several kinases, sulfotransferases and others) and proved to be selective [56].

To conclude, the PSSC approach in conjunction with natural product (olomoucine)-inspired compound library design would also have been successful in this concrete ligand design situation.

Development of nuclear hormone receptor modulators

Nuclear receptors (NRs) are ligand-inducible transcription factors consisting of a ligand-binding domain (LBD) and a DNA-binding domain (DBD). NRs are phylogenetically related proteins that have evolved through divergent evolution and are therefore grouped into a large superfamily. Structural comparison of the moderately conserved NR LBDs reveals that these domains exhibit a canonical fold, consisting of 12 α-helices, which is better conserved than the primary sequence. In the hydrophobic core of the LBD, the fully buried ligands are bound. NRs comprise receptors for hydrophobic molecules, such as steroid hormones (estrogens, glucocorticoids, progesterone, mineralocorticoids, androgens, vitamin D, ecdysone, oxysterols, bile acids and so on), retinoic acids (all-*trans* and 9-*cis* isoforms), thyroid hormones, fatty acids, leukotrienes and prostaglandins [57]. NRs are naturally switched on and off by small-molecule hormones bearing physicochemical properties that are similar to therapeutic chemical entities; therefore, they intrinsically represent attractive and promising target families in terms of therapeutic applications. Examples for the current therapeutic exploitation of NRs include the use of estrogen receptor α (ERα) antagonists (e.g. tamoxifen) for the treatment of breast cancer, and the clinical use of the structural class of thiazolidinediones (the so-called glitazones), which are agonists of peroxisome proliferator-activated receptor γ (PPARγ) and are therefore insulin sensitizers, as antidiabetic drugs [58,59].

The farnesoid X receptor (FXR) has recently been identified as a bile acid-activated NR that has a regulatory role in cholesterol metabolism. FXR controls bile-acid synthesis, conjugation and transport, as well as lipid metabolism. It is currently hypothesized that FXR senses bile acid levels and mediates the transcriptional repression of genes responsible for the conversion of excess cholesterol into bile acids, as well as the induction of genes necessary for bile acid transport. Recent advances in FXR biology suggest that FXR might represent a valuable and pharmacologically interesting target for the identification of novel drugs to treat dyslipidaemia and cholestasis [60]. However, for further

validation of FXR as a potential drug target, it is necessary to understand exactly its physiological role. A selective, cell-permeable high-affinity agonist as a tool-compound would be helpful in this context to elucidate FXR-mediated effects in a combined chemical and biological approach. To find a starting point in chemical structural space for compound library development, a clustering approach based on protein structure similarity occurring in the LBDs of the NRs would also have been successful. Thus, the LBDs of ERβ, PPARγ and FXR could be grouped into a PSSC despite low sequence homology (sequence identities amount to ~20%) (as illustrated by the superimposition of the LBDs of these NRs in Figure 5a). The natural product genistein, an isoflavone phytoestrogen (**15**; Figure 6), is found in significant levels in soy beans and soy products. Genistein binds to both ER isoforms – α and β – with moderate affinity but exhibits a preference for ERβ, acting as a partial agonist [61]. Additionally, genistein is found to be a PPARγ agonist [62]. Another known synthetic PPARγ agonist is troglitazone (**16**; Figure 6), which was in clinical use as an antidiabetic agent but was withdrawn from market as a result of its liver toxicity [63]. Genistein (**15**) and troglitazone (**16**) have in common a benzopyran core moiety. This benzopyran framework represents a privileged motif and occurs in many natural products that cover a broad spectrum of biological activities, such as antitumor, antibacterial and estrogenic effects. Thus, in light of structural conservation of the LBD fold, a compound library inspired by the structure of a natural product modulator of one member of the NR class (e.g. genistein) might also yield hits for the targeting of FXR.

Indeed, an initial screening of a combinatorial natural product-like and diversity-orientated library of 10,000 benzopyran-based small molecules, built up by Nicolaou and co-workers [64,65] using a cell-based assay for FXR activation, afforded several lead compounds (e.g. **11** and **12**; Figure 5b) with low μM activity (EC_{50} values in the range of 5–10 μM). Further elaboration of the identified lead structures following a combined solid- and solution-phase approach yielded FXR binders with EC_{50} values in the low nanomolar range, such as compound **13**, with an EC_{50} value of 188 nM, and compound **14**, with an EC_{50} value of 25 nM (Figure 5b). The benzopyran moiety in compound **14** was further deconstructed to the privileged biaryl motif [66,67]. These findings convincingly support the idea that the PSSC approach, in conjunction with natural product (genistein)-inspired compound library design, would also have been successful for the design of FXR agonists.

Further evidence for the strategy of employing the structural motif of an identified and validated ligand for one NR as rationale for the targeting of other NRs can be deduced from the cross-reactivity of particular NR ligands incorporating structurally diverse scaffolds with other NRs.

The plant sterol guggulsterone [4,17(20)-pregnadiene-3,16-dione (**17**); Figure 6], isolated from an extract of the gum resin of the guggul tree (*Commiphora mukul*), lowers low-density lipoprotein (LDL) cholesterol levels in humans. Guggulsterone (**17**) was found to be a highly efficacious antagonist of FXR, the natural ligands of which are bile acids, such as chenodeoxycholic acid [(CDCA) 3α,7α-dihydroxy-5β-cholane-24-acid; Figure 6]. Guggulsterone competes with CDCA (**18**) for binding to the LBD of FXR.

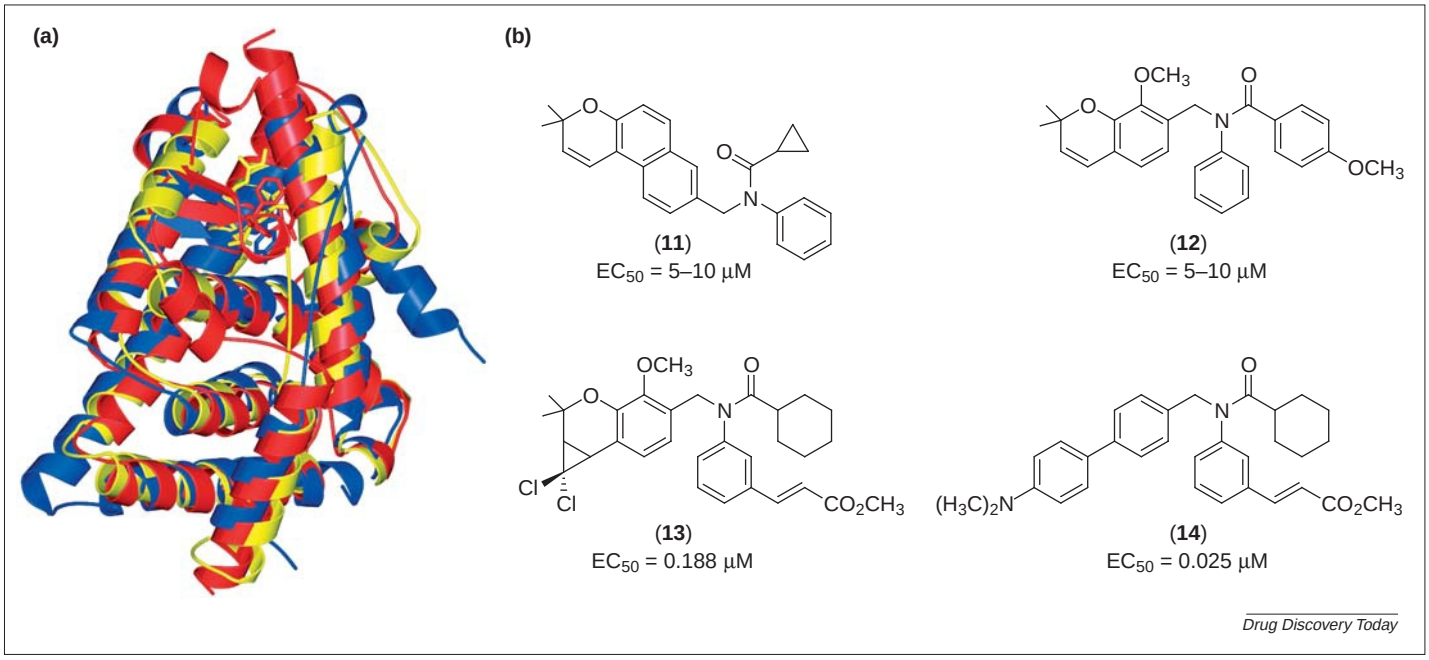

FIGURE 5

Superimposition of the X-ray structures of the ligand-binding domains of estrogen receptor β, peroxisome proliferator-activated receptor γ and farnesoid X receptor, each with bound ligand. (a) These proteins exhibit the same fold (SCOP: NR LBD) and their ligand-binding sites share a common conserved location. ERβ is shown with genistein (**15**, blue), PPARγ is complexed with rosiglitazone (red) and FXR is bound to **14** (yellow). **(b)** FXR agonists generated by combinatorial solid- and solution-phase chemistry.

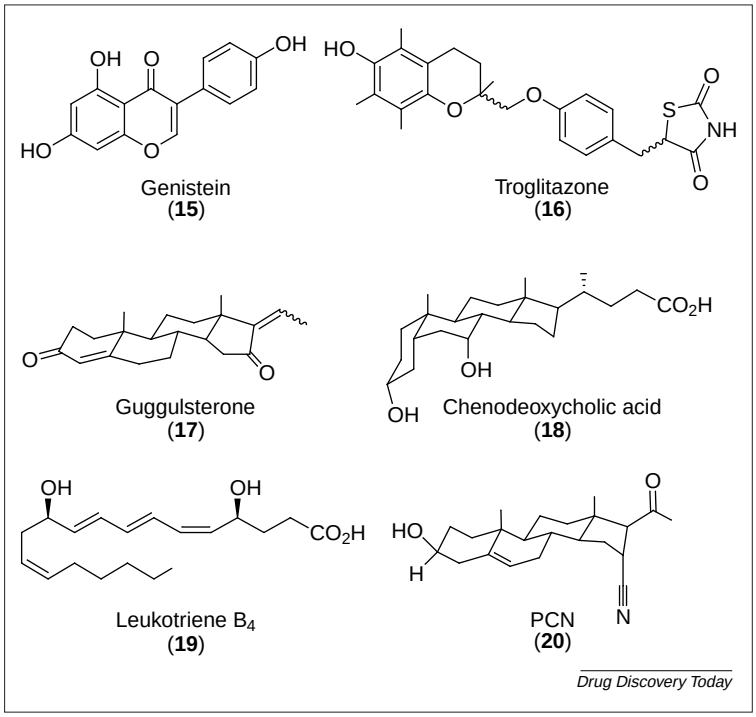

FIGURE 6

Structurally diverse natural and non-natural ligands of some nuclear hormone receptors. The structural diversity of known NR ligands suggests that NRs tolerate a wide variety of scaffolds.

It also binds to PPARα, the natural ligand of which is leukotriene B_4 (**19**; Figure 6) and to the pregnane X receptor (PXR). Guggulsterone activates PXR ~50% as effectively as the specific non-natural PXR agonist pregnenolone 16α-carbonitrile [(PCN) **20**; Figure 6)] [68]. Guggulsterone can be seen as a congener of PCN, and, therefore, its binding capability to PXR, as well as its affinity to FXR, is not surprising. By contrast, the affinity of guggulsterone to PPARα is much more unexpected when comparing their completely different chemical structures.

Comparison of the natural NR ligands led to another important finding in this context – despite their chemical diversity NR ligands have a mean molecular van der Waals volume of 318 ± 53 $Å^3$ in common. Their mean molecular weight of 368 ± 110 is less conserved. This indicates that co-evolution of receptor and ligand took place, leading to the selection of ligands with conserved volumes capable of filling the 3D space of the ligand-binding cavity in the LBD. The canonical LBD fold determines the volume of the binding pocket that the ligand must fill and thereby dictates the tolerated range of ligand volumes. Thus, molecular volumes could serve as a valuable tool for judging putative ligands [69].

The previous analysis and re-interpretation of the examples for ligand development extracted from the literature in the light of the PSSC approach convincingly suggest the applicability of the new concept to compound library design. These examples also clearly demonstrate that such ligand development can be initiated and guided successfully by the structure of a natural product that is known to bind to one member protein of a PSSC. Via refinement of the initial natural product structure, potent and selective small-molecule binders for the PSSC member proteins can be generated. These analyses encouraged us to use the PSSC approach in a real *de novo* ligand design situation. The following example represents the first successful forward application of the PSSC concept in conjunction with natural product-guided compound library development.

PSSC of Cdc25A phosphatase, acetylcholinesterase and 11β-hydroxysteroid dehydrogenases: a case study

The concept of analyzing protein domain cores, purely with respect to structural similarity without regard to functional or evolutionary arguments, and to group the respective ligand-sensing cores into a PSSC proposed by us [12,39] in conjunction with natural-product guided compound library development, was successfully applied *ab initio*, with the phosphatase Cdc25A as the initial protein of interest. A data mining and analysis strategy was developed, which enabled the identification of structurally similar protein cores from large datasets (Figure 7).

Applying this strategy, Cdc25A, acetylcholinesterase (AChE) and 11β-hydroxysteroid dehydrogenase (11βHSD) type 1 and 11βHSD2 were identified as sharing significant structural resemblance in their catalytic cores. Consequently, they were grouped into a PSSC. Despite low sequence similarity (sequence identities in the range 5–8%), the ligand-sensing cores of Cdc25A, AChE, 11βHSD1 and 11βHSD2 could be aligned structurally with root mean square deviation (RMSD) – for aligned Cα positions – values of ~3–4 Å.

These enzymes represent known or viable targets for the treatment of various diseases. Cdc25A, which regulates cell cycle progression at the G1–S checkpoint by dephosphorylating CDK2–cyclin complexes [70], could be a valuable target for the development of novel antitumor drugs. AChE hydrolyzes the neurotransmitter acetylcholine and thereby terminates impulse transmission at cholinergic synapses [71] and is currently a major target in the treatment of myasthenia gravis, glaucoma and Alzheimer's disease [72]. 11βHSD1 is essential for the local and tissue-specific activation of glucocorticoid receptors because it catalyzes the oxoreduction of cortisone to cortisol; thus, it could be a promising therapeutic target for the antagonization of glucocorticoid actions [73,74], and its inhibition is considered to be a promising approach to the treatment of obesity [75,76], the metabolic syndrome [77,78], diabetes type 2 [79,80] and cognitive dysfunction [81]. The 11βHSD2 isoenzyme catalyzes exclusively the oxidation of cortisol, and inhibition of 11βHSD2 causes sodium retention resulting in hypertension [82]. Therefore, isoenzyme-specificity is a major prerequisite for the clinical use of 11βHSD1 inhibitors.

In light of this structural similarity, a compound collection was synthesized based on a naturally occurring inhibitor of one of the enzymes. The sesterterpene dysidiolide (**21**;

Figure 8b) is an inhibitor of Cdc25A. Based on earlier investigations [14,15] and literature reports on the phosphatase-inhibiting activity of related natural products [83], it was hypothesized that the γ-hydroxybutenolide group incorporated into the natural product is a major determinant of its phosphatase inhibiting activity. Consequently, a 147-member compound collection of γ-hydroxybutenolides and closely related α,β-unsaturated five-membered lactones was synthesized and subjected to biochemical investigation for possible inhibition of Cdc25A, AChE or 11βHSD1 and 11βHSD2. Compounds displaying IC_{50} values ≤10 μM were considered as hits (Figure 8b). Of the 147 compounds investigated, 42 qualified as hits in the Cdc25A assay. The most potent compound (**22**) had an IC_{50} value of 350 nM, which is significantly lower than the reported IC_{50} value for dysidiolide (9.4 μM [84]). Three compounds inhibited AChE with IC_{50} values of 1.3–4.5 μM. The collection contained three 11βHSD1 inhibitors with IC_{50} values of 7.8–10.0 μM and four 11βHSD2 inhibitors with IC_{50} values of 2.4–6.7 μM. Thus, the hit rates for the enzymes identified as being similar to Cdc25A are ~2–3%.

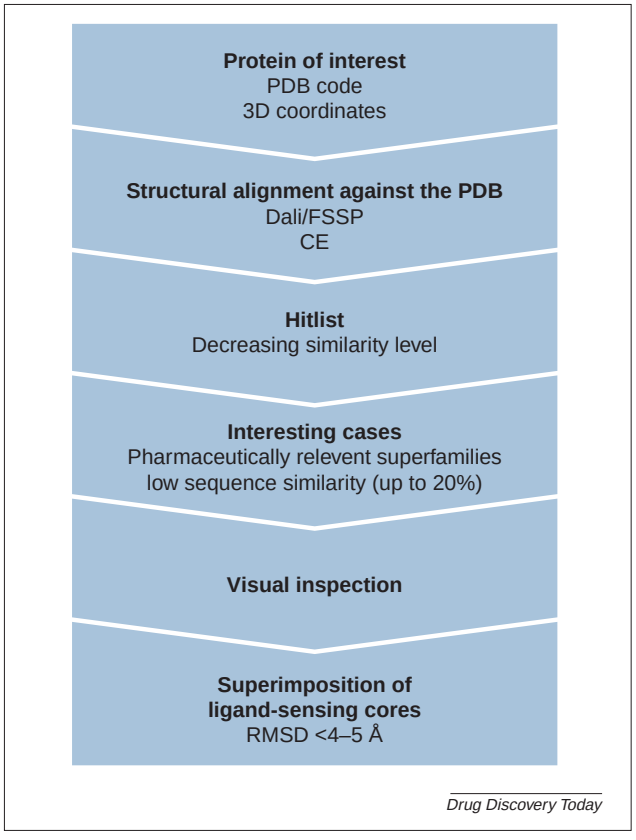

FIGURE 7

Database search strategy and procedure developed for the identification of protein structure similarity clusters. Database searches, for example, in the Dali/FSSP [89,90] (http://www.ebi.ac.uk/dali) and the Combinatorial Extension (CE) [91] (http://cl.sdsc.edu/ce.html) databases, using the 3D coordinates of a query protein could provide insights into their structural neighborhood. For compound library development, the ligand-sensing cores of the proteins are of paramount importance, therefore, it is imperative that these relevant parts of the protein domains share structural similarity.

Even at this comparably small library size, the hits displayed a pronounced degree of selectivity for individual enzymes and also for the isoenzymes 11βHSD1 and 11βHSD2. Thus, compound **22** was a significantly more potent inhibitor for Cdc25A than for the other enzymes. Most remarkably, the α,β-unsaturated lactone **23** inhibited only the therapeutically relevant 11βHSD1 but not, or only weakly, the other enzymes investigated. In addition, a furan derivative (**24**) was identified as an inhibitor for Cdc25A and 11βHSD2. However, a selective inhibitor for AChE could not be discovered.

This example gives evidence for the *de novo* applicability of the PSSC concept. It clearly demonstrates that 3D protein structures generated using homology modeling techniques can also be considered. However, it is expected that the reliability of the PSSC approach will increase with the growing number of experimentally determined protein structures becoming available.

Conclusion: a new guiding principle for chemical genomics?

The examples detailed here demonstrate that the PSSC approach can serve as a conceptually new principle guiding the development of compound libraries, in particular for medicinal chemistry research. However, beyond this, the use of PSSCs could open up new opportunities for research in the currently developing field of 'chemical genomics'. In a general sense, chemical genomics can be defined as the genomic response to chemical compounds, in other words, chemistry is employed to probe a biological system. A more focused, workable definition appears to be the identification of small-molecule lead-like compounds for a member of a gene-family product and the subsequent use of these compounds to elucidate the function of other (disease-associated) members of the gene family. Currently, in this approach, the gene family products are predominantly classified on the basis of sequence similarities and function, for example, into kinases, phosphatases and proteases (Figure 9).

A protein domain core-centered approach that considers domain organization and architecture, however, could provide a new guiding principle for the combinatorial development of compounds that will pave the way for a new series of chemical proteomics and genomics experiments. Accordingly, a family of gene products (proteins) of interest would be dissected in structural terms, for example, into domains. After domain assignment, structural comparison of the ligand-sensing cores with known domains and/or folds would be performed, leading to a cluster of structurally related domain cores that might share little sequence homology (Figure 9). This pool of structurally similar ligand-sensing cores with their respective ligands could be used for the generation of potent and selective small-molecule modulators of protein function of the PSSC member proteins. The structures of known ligands for a spatially similar reference domain core, constitute biologically

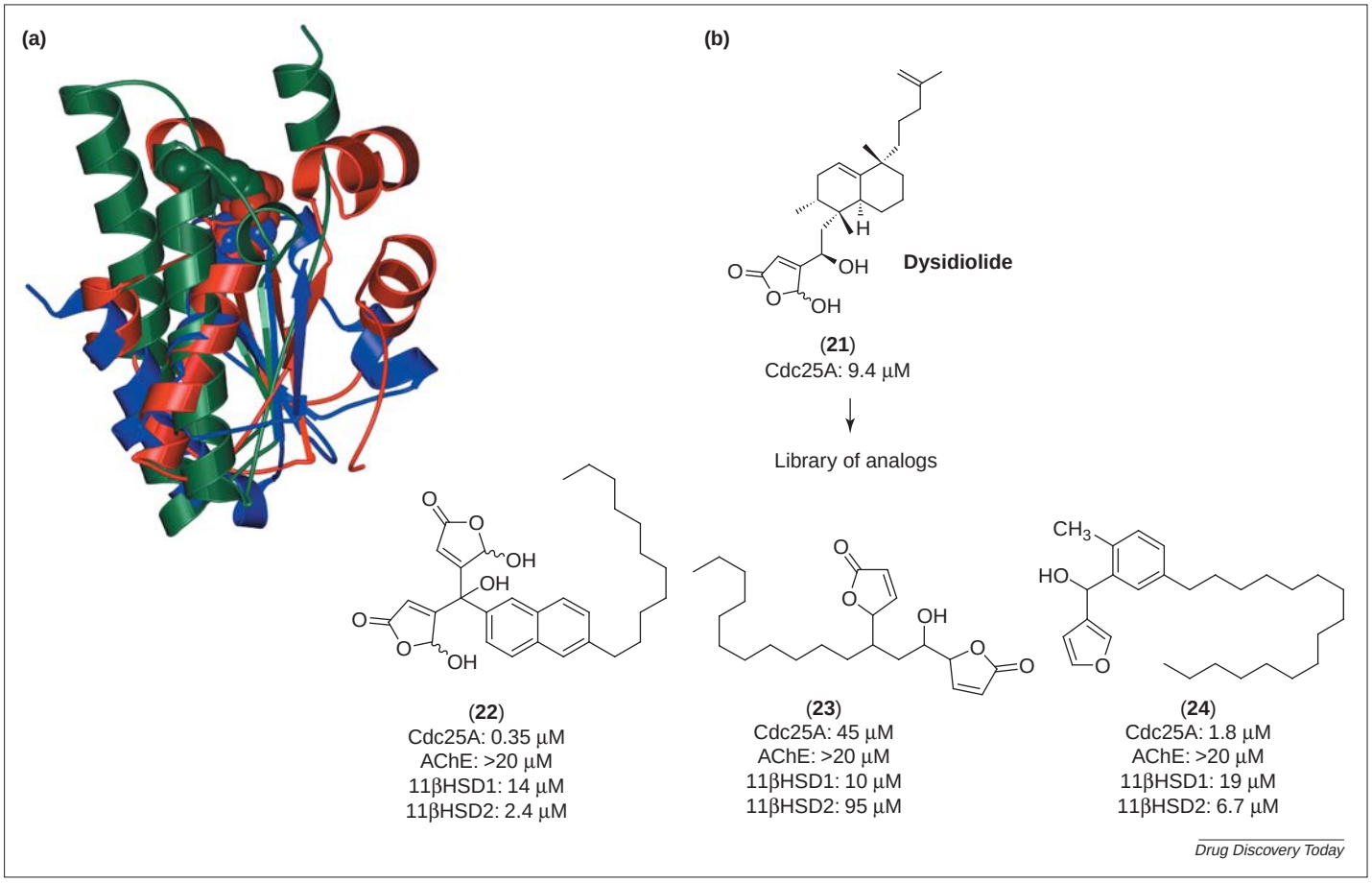

FIGURE 8

Superimposition of the catalytic cores of Cdc25A, 11β-hydroxysteroid dehydrogenase 1 and acetylcholinesterase. (a) The key catalytic residues, Cys430 (Cdc25A; red), Tyr183 (11βHSD1; green) and Ser200 (AChE; blue), shown in Corey–Pauling–Koltun representation, are located similarly. **(b)** Analogs of the naturally occurring Cdc25A inhibitor dysidiolide profiled against the PSSC member proteins Cdc25A, AChE, 11βHSD1 and 11βHSD2 (IC_{50} values are given).

validated starting points in chemical structural space for the design of focused libraries yielding comparably high hit rates when screened against the PSSC member proteins. Selectivity and enhanced potency can be achieved by generating diversity around the small-molecule binder core structure, thus taking into account the requirements of the individual binding pockets, possibly harboring substantially different amino acid residues. This strategy initially reduces complexity and focuses on the 3D similarity of protein domain cores, and ultimately leads to structural frameworks guiding compound library development.

Non-natural substances or natural products known to bind to one member protein of the PSSC could serve as leitmotifs (i.e. for compound library development). For natural products, we postulate that their evolutionarily selected scaffolds represent biologically pre-validated structures providing basic affinity to the protein domain cores, with which they were evolved to interact. Natural products can thus be regarded as inherently promising guiding compounds for the design of domain selective small-molecule modulators of protein function. Of course, this prerequisite is not only fulfilled by natural products. Non-natural synthetic small-molecule ligands with known biological relevance can also be regarded as valuable starting points. In a sense, their binding properties have been evolved in the course of an accelerated artificial evolutionary process.

Once a biologically relevant structural framework has been found, the varying requirements of the different binding sites can be addressed by generating diversity around this core structure, thus enabling evolution of potent and selective binders. This can be accomplished using a library approach that can be supported by molecular modeling techniques. Often, the synthesis of natural product-derived compound libraries requires the development of demanding multi-step synthesis sequences, including, for example, enantioselective transformations. Thus, the initial investment in the synthesis of such libraries might be high. However, compound libraries generated following this approach can be small compared with classical combinatorial libraries primarily developed on the basis of chemical feasibility and accessibility. Thus, a higher developmental investment for the generation of such focused libraries, as a result of the structural complexity of the natural products used as sources of inspiration, is

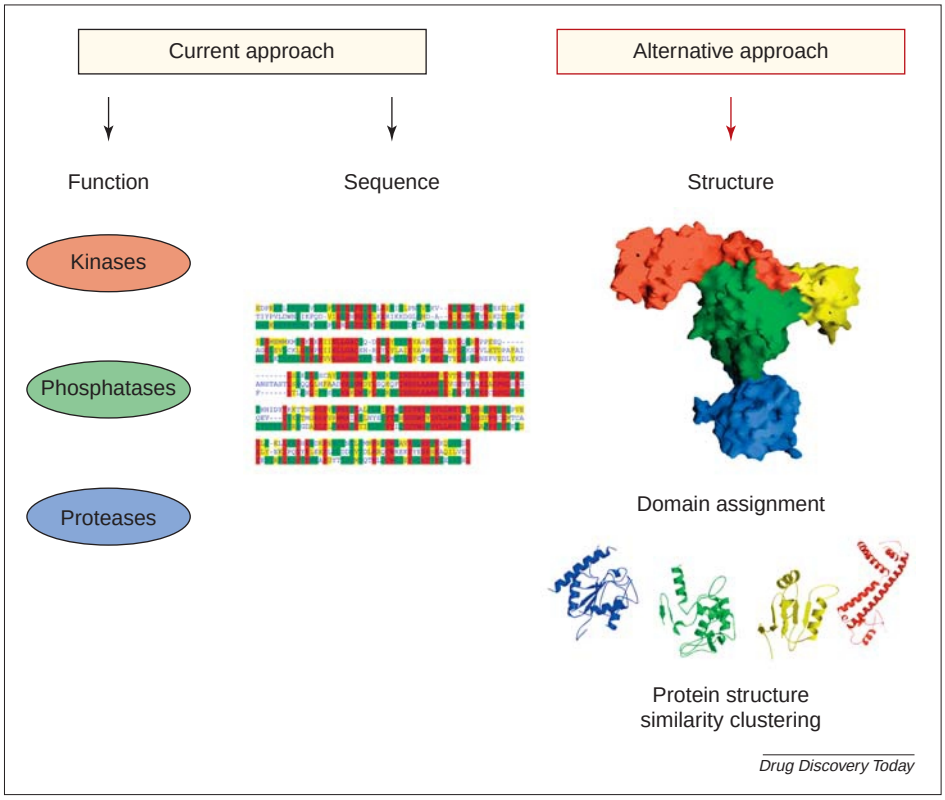

FIGURE 9

Approaches for protein categorization. The currently predominating approach in chemical genomics, which is based on the clustering of target proteins according to their sequence and function, might be complemented by an alternative approach based on a purely structural view of protein domains or cores.

justified because such libraries will yield high hit rates and – most importantly – biologically pre-validated hits. In the light of this argument, further development of chemical methods enabling the rapid and efficient synthesis of complex molecules in library format has to be regarded as being of utmost importance [10,14,15,85–88].

To summarize, in the initial step of the PSSC approach, the overall structural resemblance of protein domain cores is used as the guiding principle in choosing possible small-molecule binder scaffolds. In a second step, the structural diversity occurring in the binding sites of the PSSC member proteins is addressed by synthesizing a compound library. Thereby, chemical entities are identified that efficiently address the biological diversity found in the binding sites, thus yielding selective and potent binders. The advantage of an initial 'indeterminateness' when comparing overall domain core structures is that predicted and modeled protein structures with a particular tolerance with respect to the binding sites can also be considered because, finally, the indeterminateness is overcome by the combinatorial approach.

The PSSC concept should be particularly helpful in the initial stages of compound development and screening when little might be known about the function of a newly discovered protein. Thus, a PSSC strategy serves as an abstracting principle that enables the identification of novel compound classes for a given target. The identified structural frameworks then have to be refined in a medicinal chemistry program to optimize selectivity and to reduce unwanted activities.

Acknowledgements

We thank Ingrid Vetter (Max Planck Institute of Molecular Physiology, Dortmund) for continuing stimulating discussions. This work was supported by the Max-Planck-Gesellschaft, the Deutsche Forschungsgemeinschaft and the Studienstiftung des deutschen Volkes (doctoral scholarship for M.A.K.).

References

1 Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969
2 Golebiowski, A. *et al.* (2001) Lead compounds discovered from libraries. *Curr. Opin. Chem. Biol.* 5, 273–284
3 Mason, J.S. and Hermsmeier, N.A. (1999) Diversity assessment. *Curr. Opin. Chem. Biol.* 3, 342–349
4 Spencer, R.W. (1998) High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Biotechnol. Bioeng.* 61, 61–67
5 Walters, W.P. *et al.* (1999) Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* 3, 384–387
6 Ajay, A. *et al.* (1998) Can we learn to distinguish between 'drug-like' and 'nondrug-like' molecules? *J. Med. Chem.* 41, 3314–3324
7 Sadowski, J. and Kubinyi, H. (1998) A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329
8 Ghose, A.K. *et al.* (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1, 55–68
9 Lee, M.L. and Schneider, G. (2001) Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* 3, 284–289
10 Breinbauer, R. *et al.* (2002) From protein domains to drug candidates – natural products as guiding principles in the design and synthesis of compound libraries. *Angew. Chem. Int. Ed. Engl.* 41, 2879–2890
11 Koch, M.A. *et al.* (2003) Protein structure similarity as guiding principle for combinatorial library design. *Biol. Chem.* 384, 1265–1272
12 Koch, M.A. and Waldmann, H. (2004) Natural Product-derived Compound Libraries and Protein Structure Similarity as Guiding Principles for the Discovery of Drug Candidates. In *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective* (Kubinyi, H. and Müller, G., eds), pp. 377–403, Wiley-VCH
13 Evans, B.E. *et al.* (1988) Methods for drug discovery – development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* 31, 2235–2246
14 Brohm, D. *et al.* (2002) Natural products are biologically validated starting points in structural space for compound library development: solid-phase synthesis of dysidiolide-derived phosphatase inhibitors. *Angew. Chem. Int. Ed. Engl.* 41, 307–311
15 Brohm, D. *et al.* (2002) Solid-phase synthesis of dysidiolide-derived protein phosphatase inhibitors. *J. Am. Chem. Soc.* 124, 13171–13178
16 Grishin, N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.* 134, 167–185
17 Lupas, A.N. *et al.* (2001) On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J. Struct. Biol.* 134, 191–203
18 Govindarajan, S. *et al.* (1999) Estimating the total number of protein folds. *Proteins* 35, 408–414

19 Ponting, C.P. *et al.* (2000) Evolution of Domain Families. In *Advances in Protein Chemistry* (Vol. 54), pp. 185–244
20 Murzin, A.G. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540
21 Andreeva, A. *et al.* (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.* 32, D226–D229
22 Orengo, C.A. *et al.* (1994) Protein superfamilies and domain superfolds. *Nature* 372, 631–634
23 Chothia, C. (1992) Proteins – 1000 families for the molecular biologist. *Nature* 357, 543–544
24 Zhang, C.T. (1997) Relations of the numbers of protein sequences, families and folds. *Protein Eng.* 10, 757–761
25 Wang, Z.X. (1998) A re-estimation for the total numbers of protein folds and superfamilies. *Protein Eng.* 11, 621–626
26 Zhang, C.O. and DeLisi, C. (1998) Estimating the number of protein folds. *J. Mol. Biol.* 284, 1301–1305
27 Wolf, Y.I. *et al.* (2000) Estimating the number of protein folds and families from complete genome data. *J. Mol. Biol.* 299, 897–905
28 Coulson, A.F.W. and Moult, J. (2002) A unifold, mesofold, and superfold model of protein fold use. *Proteins* 46, 61–71
29 Koonin, E.V. *et al.* (2002) The structure of the protein universe and genome evolution. *Nature* 420, 218–223
30 Weir, M. *et al.* (2001) Insights into protein function through large-scale computational analysis of sequence and structure. *Trends Biotechnol.* 19, S61–S66
31 Danchin, A. (1999) From protein sequence to function. *Curr. Opin. Struct. Biol.* 9, 363–367
32 Tatusov, R.L. *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28
33 Anantharaman, V. *et al.* (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* 7, 12–20
34 Gerlt, J.A. and Babbitt, P.C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* 70, 209–246
35 Todd, A.E. *et al.* (1999) Evolution of protein function, from a structural perspective. *Curr. Opin. Chem. Biol.* 3, 548–556
36 Frye, S.V. (1999) Structure–activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* 6, R3–R7
37 Jacoby, E. *et al.* (2003) Chemogenomics knowledge-based strategies in drug discovery. *Drug News Perspect.* 16, 93–102
38 Russell, R.B. *et al.* (1998) Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* 282, 903–918
39 Koch, M.A. *et al.* (2004) Compound library development guided by protein structure similarity clustering and natural product structure. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16721–16726
40 Thunnissen, M.M.G.M. *et al.* (2001) Crystal structure of human leukotriene A4 hydrolase, a bifunctional enzyme in inflammation. *Nat. Struct. Biol.* 8, 131–135
41 Rudberg, P.C. *et al.* (2004) Leukotriene A4 hydrolase: identification of a common carboxylate recognition site for the epoxide hydrolase and aminopeptidase substrates. *J. Biol. Chem.* 279, 27376–27382
42 Orning, L. *et al.* (1991) Leukotriene A4 hydrolase – inhibition by bestatin and intrinsic aminopeptidase activity establish its functional resemblance to metallohydrolase enzymes. *J. Biol. Chem.* 266, 1375–1378
43 Yuan, W. *et al.* (1993) Development of selective tight-binding inhibitors of leukotriene A4 hydrolase. *J. Med. Chem.* 36, 211–220
44 Ollmann, I.R. *et al.* (1995) Investigation of the inhibition of leukotriene A4 hydrolase. *Bioorg. Med. Chem.* 3, 969–995
45 Hogg, J.H. *et al.* (1995) Amino hydroxamic acids as potent inhibitors of leukotriene A4 hydrolase. *Bioorg. Med. Chem.* 3, 1405–1415
46 Zhang, M.Q. (1997) Leukotriene A4 hydrolase: a new target for anti-inflammatory intervention. *Curr. Med. Chem.* 4, 67–78
47 Thunnissen, M.M.G.M. *et al.* (2002) Crystal structures of leukotriene A4 hydrolase in complex with captopril and two competitive tight-binding inhibitors. *FASEB J.* 16, 1648–1650
48 Norman, T.C. *et al.* (1996) A structure-based library approach to kinase inhibitors. *J. Am. Chem. Soc.* 118, 7430–7431
49 Gray, N.S. *et al.* (1998) Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 281, 533–538
50 Chang, Y.T. *et al.* (1999) Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem. Biol.* 6, 361–375
51 Chang, Y.T. *et al.* (2002) Purine-based inhibitors of inositol-1,4,5-trisphosphate-3-kinase. *Chembiochem* 3, 897–901
52 Kakuta, Y. *et al.* (1997) Crystal structure of estrogen sulphotransferase. *Nat. Struct. Biol.* 4, 904–908
53 Kakuta, Y. *et al.* (1998) The sulfuryl transfer mechanism – crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *J. Biol. Chem.* 273, 27325–27330
54 Armstrong, J.I. *et al.* (2000) Discovery of carbohydrate sulfotransferase inhibitors from a kinase-directed library. *Angew. Chem. Int. Ed. Engl.* 39, 1303–1306
55 Verdugo, D.E. *et al.* (2001) Discovery of estrogen sulfotransferase inhibitors from a purine library screen. *J. Med. Chem.* 44, 2683–2686
56 Chapman, E. *et al.* (2002) A potent and highly selective sulfotransferase inhibitor. *J. Am. Chem. Soc.* 124, 14524–14525
57 Robinson-Rechavi, M. *et al.* (2003) The nuclear receptor superfamily. *J. Cell Sci.* 116, 585–586
58 Schapira, M. (2002) Pharmacogenomics opportunities in nuclear receptor targeted cancer therapy. *Curr. Cancer Drug Targets* 2, 243–256
59 Schapira, M. *et al.* (2000) Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1008–1013
60 Claudel, T. *et al.* (2004) The farnesoid X receptor: a novel drug target? *Expert Opin. Investig. Drugs* 13, 1135–1148
61 Pike, A.C.W. *et al.* (1999) Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *EMBO J.* 18, 4608–4618
62 Dang, Z.C. *et al.* (2003) Peroxisome proliferator-activated receptor gamma (PPAR gamma) as a molecular target for the soy phytoestrogen genistein. *J. Biol. Chem.* 278, 962–967
63 Van Gaal, L. and Scheen, A.J. (2002) Are all glitazones the same? *Diabetes Metab. Res. Rev.* 18 (Suppl. 2), S1–S4
64 Nicolaou, K.C. *et al.* (2000) Natural product-like combinatorial libraries based on privileged structures. 1. General principles and solid-phase synthesis of benzopyrans. *J. Am. Chem. Soc.* 122, 9939–9953
65 Nicolaou, K.C. *et al.* (2000) Natural product-like combinatorial libraries based on privileged structures. 2. Construction of a 10 000-membered benzopyran library by directed split-and-pool chemistry using NanoKans and optical encoding. *J. Am. Chem. Soc.* 122, 9954–9967
66 Nicolaou, K.C. *et al.* (2003) Discovery and optimization of non-steroidal FXR agonists from natural product-like libraries. *Org. Biomol. Chem.* 1, 908–920
67 Downes, M. *et al.* (2003) A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell* 11, 1079–1092
68 Urizar, N.L. *et al.* (2002) A natural product that lowers cholesterol as an antagonist ligand for FXR. *Science* 296, 1703–1706
69 Bogan, A.A. *et al.* (1998) Natural ligands of nuclear receptors have conserved volumes. *Nat. Struct. Biol.* 5, 679–681
70 Fauman, E.B. *et al.* (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* 93, 617–625
71 Sussman, J.L. *et al.* (1991) Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* 253, 872–879
72 Ibach, B. and Haen, E. (2004) Acetylcholinesterase inhibition in Alzheimer's disease. *Curr. Pharm. Des.* 10, 231–251
73 Walker, B.R. and Seckl, J.R. (2003) 11beta-hydroxysteroid dehydrogenase type 1 as a novel therapeutic target in metabolic and neurodegenerative disease. *Expert Opin. Ther. Targets* 7, 771–783
74 Chrousos, G.P. (2004) Is 11beta-hydroxysteroid dehydrogenase type 1 a good therapeutic target for blockade of glucocorticoid actions? *Proc. Natl. Acad. Sci. U. S. A.* 101, 6329–6330
75 Schweizer, R.A. *et al.* (2003) A rapid screening assay for inhibitors of 11beta-hydroxysteroid dehydrogenases (11beta-HSD): flavanone selectively inhibits 11beta-HSD1 reductase activity. *Mol. Cell. Endocrinol.* 212, 41–49
76 Masuzaki, H. *et al.* (2001) A transgenic model of visceral obesity and the metabolic syndrome. *Science* 294, 2166–2170
77 Masuzaki, H. and Flier, J.S. (2003) Tissue-specific glucocorticoid reactivating enzyme, 11beta-hydroxysteroid dehydrogenase type 1 (11beta-HSD1) – a promising drug target for the treatment of metabolic syndrome. *Curr. Drug Targets Immune Endocr. Metabol. Disord.* 3, 255–262
78 Paterson, J.M. *et al.* (2004) Metabolic syndrome without obesity: hepatic overexpression of 11beta-hydroxysteroid dehydrogenase type 1 in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7088–7093
79 Alberts, P. *et al.* (2002) Selective inhibition of 11beta-hydroxysteroid dehydrogenase type 1 decreases blood glucose concentrations in hyperglycaemic mice. *Diabetologia* 45, 1528–1532

80 Ross, S.A. *et al.* (2004) Chemistry and biochemistry of type 2 diabetes. *Chem. Rev.* 104, 1255–1282
81 Sandeep, T.C. *et al.* (2004) 11Beta-hydroxysteroid dehydrogenase inhibition improves cognitive function in healthy elderly men and type 2 diabetics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6734–6739
82 New, M.I. and Wilson, R.C. (1999) Steroid disorders in children: congenital adrenal hyperplasia and apparent mineralocorticoid excess. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12790–12797
83 Lyon, M.A. *et al.* (2002) Dual-specificity phosphatases as targets for antineoplastic agents. *Nat. Rev. Drug Discov.* 1, 961–976
84 Gunasekera, S.P. *et al.* (1996) Dysidiolide: a novel protein phosphatase inhibitor from the Caribbean sponge *Dysidea etheria* de Laubenfels. *J. Am. Chem. Soc.* 118, 8759–8760
85 Abreu, P.M. and Branco, P.S. (2003) Natural product-like combinatorial libraries. *J. Braz. Chem. Soc.* 14, 675–712
86 Arya, P. *et al.* (2002) Toward high-throughput synthesis of complex natural product-like compounds in the genomics and proteomics age. *Chem. Biol.* 9, 145–156
87 Nicolaou, K.C. and Pfefferkorn, J.A. (2001) Solid-phase synthesis of complex natural products and libraries thereof. *Biopolymers* 60, 171–193
88 Barun, O. *et al.* (2004) Asymmetric solid-phase synthesis of 6,6-spiroketals. *Angew. Chem. Int. Ed. Engl.* 43, 3195–3199
89 Holm, L. and Sander, C. (1996) The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.* 24, 206–209
90 Holm, L. and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* 25, 231–234
91 Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11, 739–747