J Comput Aided Mol Des (2008) 22:761–772
DOI 10.1007/s10822-008-9221-8

CONFIRM: connecting fragments found in receptor molecules

David C. Thompson · R. Aldrin Denny · Ramaswamy Nilakantan · Christine Humblet · Diane Joseph-McCarthy · Eric Feyfant

Received: 19 December 2007 / Accepted: 17 May 2008 / Published online: 9 July 2008

Abstract A novel algorithm for the connecting of fragment molecules is presented and validated for a number of test systems. Within the CONFIRM (Connecting Fragments Found in Receptor Molecules) approach a pre-prepared library of bridges is searched to extract those which match a search criterion derived from known experimental or computational binding information about fragment molecules within a target binding site. The resulting bridge 'hits' are then connected, in an automated fashion, to the fragments and docked into the target receptor. Docking poses are assessed in terms of root-mean-squared deviation from the known positions of the fragment molecules, as well as docking score should known inhibitors be available. The creation of the bridge library, the full details and novelty of the CONFIRM algorithm, and the general applicability of this approach within the field of fragment-based de novo drug design are discussed.

Keywords Bridge library · Connecting fragments · Fragment-based de novo design · Molecular docking · Structure-based drug design

Introduction

Structure-based drug design is a powerful technique widely employed within the pharmaceutical and biotechnology fields whenever structural information is available. Initial lead compounds are optimized through an iterative process involving binding mode prediction, computational binding assessment, chemical synthesis of proposed compounds, and biological testing. The success of such a cycle is influenced both by the source of the initial lead compound and its quality.

There are a number of ways of generating an initial lead. In a conventional high-throughput screening (HTS) campaign large libraries of up to a million compounds are screened in the hopes of finding relatively potent leads; those with IC_{50} values less than 10 μM. While this method has been shown to be useful [1, 2], it is intrinsically expensive and resource intensive. Virtual high-throughput screening (vHTS) can be a useful complement to a standard HTS campaign both to identify false negatives in the experimental screen of the corporate collection and to screen additional libraries of compounds. Virtual screening methodologies have the benefit that they are relatively inexpensive—both in time and resources—and can be run routinely [3–5]. Fragment screening is yet another approach [6].

The formal concepts underpinning the fragment approach date back to the work of Jencks [7] and Ariens [8] who demonstrated that drug-like molecules can be considered as combinations of two or more individual binding epitopes (or fragments). This is a rationalization of something that the medicinal chemistry community regularly utilizes; the entire field of pharmacophore analysis is based on the notion of 'fragmenting' a molecule into biologically relevant features.

A fragment molecule is typically small (molecular weight (MW) 120–300, with between 8 and 18 non-hydrogen atoms) and has less functionality than most lead-like molecules. Therefore, fragment hits are correspondingly weaker

D. C. Thompson · R. Aldrin Denny · R. Nilakantan · D. Joseph-McCarthy · E. Feyfant (✉)
Wyeth Research, Chemical and Screening Sciences, 200 CambridgePark Drive, Cambridge, MA 02140, USA
e-mail: efeyfant@wyeth.com

C. Humblet
Wyeth Research, Chemical and Screening Sciences, 865 Ridge Road, Princeton, NJ 08543, USA

binders than most hits from an HTS; typical binding affinities are in the high μM to mM range. Fragment molecules ideally have a higher number of atoms which are directly involved in the binding event and are thus described as efficient binders. [9–11]. Given this level of binding affinity, suitable biophysical screening methods include Nuclear Magnetic Resonance (NMR) [12, 13] and X-ray crystallography [14, 15], both of which afford important structural information. The 'SAR by NMR' approach has been used to link fragments into a single molecular entity [13]; specific examples include its application to FK506-binding protein [13] and matrix metalloproteinases [16]. Fragment-based screening approaches such as NMR and high-throughput X-ray crystallography can help to identify novel chemotypes by suggesting starting orientations of molecular building blocks for subsequent computational linking or growing. Subsequent chemical optimization of low MW fragments is likely to produce leads whose MW is still within the range desired for lead-likeness [17]. A number of reviews of fragment-based lead discovery have recently been published [18–21].

The optimization of a fragment hit can also be guided through the use of computational methods and the many computational strategies can be broadly characterized as either growing or linking. The use of computational approaches to the fragment-based de novo design problem has been recently reviewed [20].

Initial fragments can thus be identified using NMR or X-ray [22–24] experimental fragment-binding information, docking software, or determination of hot-spots within the protein active site using computational tools such as HSITE [25], HIPPO [26], or MCSS [27, 28]

In the growing approach an initial fragment is progressively expanded in an attempt to improve the interactions between the receptor and the ligand. There are numerous examples of software applying the growing approach, e.g., SPROUT [29, 30], LEGEND [31], LUDI [32], GROWMOL [33], LigBuilder [34] SkelGen [35, 36], and SMoG [37].

In the linking approach, fragment building blocks are positioned in the target binding site as described above and then automatically connected to each other by linkers to yield a complete molecule which satisfies all of the key interactions. The tacit assumption here is that the binding affinity of fragments is additive and, moreover, that the affinity contribution from the linker is negligible, or favorable, and that the loss in rigid-body entropy on binding of all of the components of the molecule is small [38]. LUDI [32], HOOK [39], PRO_LIGAND [40], LigBuilder [34], SPLICE/RACHEL [41] and CAVEAT [42, 43] are examples of software using the linking approach.

The novel algorithm described here, CONFIRM, is a linking approach and is most similar to CAVEAT. With CAVEAT, the orientation of the bonds to substituents is retained to form a specific spatial relationship used to search three-dimensional databases to identify molecular scaffolds that can maintain the 3D orientation of the substituents. CAVEAT was originally designed to develop confomationally constrained peptides and peptidomimetics [44–46], but has also been used for drug design [47, 48]. In contrast CONFIRM retrieves molecular fragments using interatomic distances and atom types. Specifically, it searches a database of bridges using a fast substructure pattern search function to identify matching 'hits', which are then linked to pre-defined fragment molecules in a combinatorial fashion. Candidate molecules are subsequently docked into the target binding site and the resulting docking poses analyzed.

This novel linking strategy has been developed for a number of reasons. Firstly, the Wyeth corporate database has previously been mined resulting in the creation of a bridge fragments database [49]; the CONFIRM approach came naturally to leverage this chemical equity which is generally applicable to the problem of molecular design. Secondly, the use of automated workflow tools for the manipulation of chemical data enabled the rapid deployment of a sophisticated protocol which could be delivered through the desktop, behind a web-based interface. Such a deployment permits a high degree of customization, so project specific requirements can be easily addressed.

In the succeeding sections of this paper the CONFIRM algorithm is described in detail. Various case studies are presented showing successful application of the methodology. The examples have been retrieved from the Protein Data Bank and each present two well resolved binding pockets containing distinct fragments linked by a bridge. Measures of success for the method include the ability to identify a matching bridge in the library, and further to ensure that this bridge can lead to a fully connected ligand that retains the binding mode of the original fragments. The applicability of this method to the problem of fragment-based drug design is discussed.

Methodology

We describe below the preparation of the bridge library database, and the steps involved in the CONFIRM algorithm: Search, Connect, Prepare, and Dock. While we present the specific computational tools used in this implementation of CONFIRM, any similar computational technique for docking, or ligand preparation could be readily substituted. Indeed, the CONFIRM algorithm has been implemented into a single protocol using the Pipeline Pilot workflow tool [50] which efficiently integrates the modular components of the algorithm.

Bridge library

The library has been derived from our corporate database using the ring scaffold descriptors previously described by Nilakantan et al [see 49 and references therein]. Specifically, a 'ring scaffold' is derived by deleting all acyclic single-bonded appendages on ring systems and linkers connecting ring systems [51, 52]. A 'bridge' is extracted from any given ring scaffold by removing pairs of connected rings while retaining *the ring anchor atoms* as described in Fig. 1. Inclusion of the anchor atoms provides the implicit inclusion of some degree of synthetic accessibility, a point of great importance for de novo ligand design [53]. The final library contained 8951 unique bridges which were further filtered as follows:

1. Valence check.
2. Restrict to carbon, nitrogen, sulfur, oxygen, and phosphorus atoms.
3. ≤ 3 or ≤ 4 rotatable bonds.
4. Molecular weight ≤ 200.
5. Remove bridges containing terminal double bonds as illustrated in Fig. 2a.

Finally, for each bridge, a single three-dimensional conformer is generated and the entire library is exported as an SD file. Four libraries were generated using rotatable bond cutoffs of 3 and 4, with and without conformational expansion. Conformational expansion was performed using OMEGA2.0 [54]; default parameters were used, excepting the ewindow parameter which was set to 5 kcal/mol. The resulting bridge libraries used in this work are summarized in Table 1.

Figures 3 and 4 illustrate the distribution profiles with respect to number of rotatable bonds and molecular weight, for the un-expanded libraries containing 935 and 2090 bridges, respectively. For each of the four libraries, a

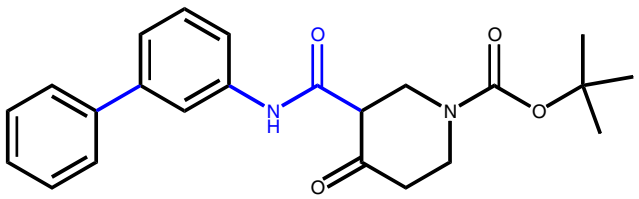


Fig. 1 Sample molecule showing bridges as defined in reference [49], these are highlighted in blue

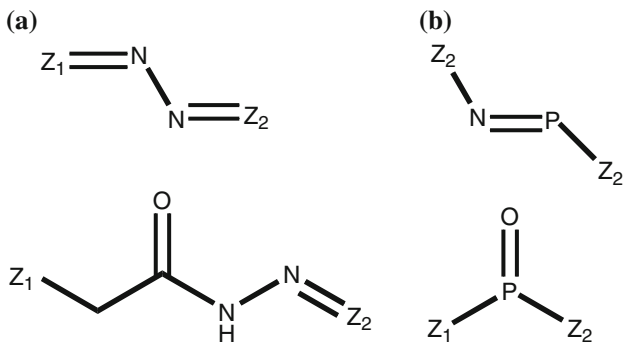


Fig. 2 Examples of removed bridges during (**a**) the database preparation stage, and (**b**) during the LigPrep preparation stage

Table 1 Effect of rotatable bond filter, and conformational expansion on the size of the bridge library

Rotatable bond filter	Un-expanded no. of bridge entries	OMEGA2.0 Expanded no. of bridge entries
≤ 3	935[a]	3407[c]
≤ 4	2090[b]	13388[d]

[a] Library Lib3
[b] Library Lib4
[c] Library Lib3E
[d] Library Lib4E

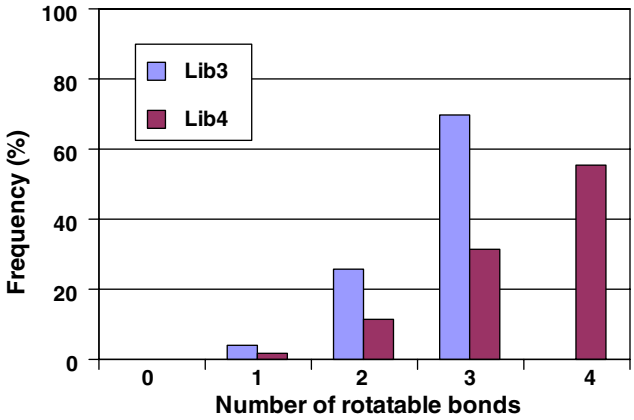


Fig. 3 Rotatable bond distribution, for the bridge libraries Lib3 and Lib4 containing 935 and 2090 unique entities, respectively

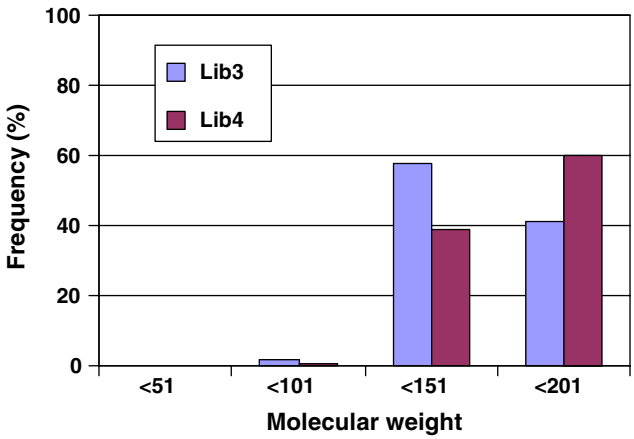


Fig. 4 Molecular weight distribution, for the bridge libraries Lib3 and Lib4 containing 935 and 2090 unique entities, respectively

terminal atom distance analysis was performed to investigate the 'length' distribution of the bridges (Fig. 5). The libraries themselves are diverse, with a mean pair-wise similarity of ~0.3 for both Lib3 and Lib4 as measured using MACSS fingerprints and the Tanimoto coefficient within MOE [55].

While this initial 'proof-of-principle' approach focuses on the use of libraries of simple acyclic bridge connectors, any chemical functionality could be used to create other connecting bridges. Work is currently progressing to extend the libraries to include diverse functional groups; including ring systems of various sizes. Given the modular nature of the algorithm such changes are straightforward, yet afford flexibility and customization in how CONFIRM is used.

Search

Given a set of fragment molecules, the distances between the attachment points and their atom types are extracted from the structure file, and form a query used to search the bridge libraries. An attachment point is defined by the user and can be any atom on either of the fragment molecules. The atom types of the attachment points, along with the distance between them, form the search criterion used with the in-house software PharmFinder [Cole B, Manas ES (unpublished work)]. PharmFinder can take multiple input database formats, and is very fast—typically taking fractions of a second to perform a search against any one of the bridge libraries. PharmFinder first requires the definition of the features of interest or the attachment points as defined by a substructure pattern. We have used the SMARTS language to describe molecular patterns and properties [56]. PharmFinder also requires the distance between features in Angstroms. Given the form of our databases we have used the following as the distance query between attachment points: $D = \lceil d \rceil \pm \text{tol}$ [Å], where d is the measured distance between attachment points and $\lceil\ \rceil$ is the ceiling operator

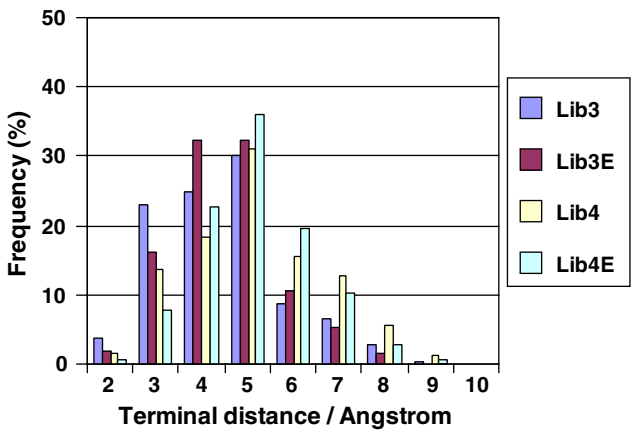


Fig. 5 Terminal distance distribution of the four bridge libraries detailed in Table 1

(the operator which gives the smallest integer $\geq d$). A distance tolerance (tol) of 1.4 Å has been used in this study and has been determined empirically. When using CONFIRM, the tolerance accounts for the fact that PharmFinder searches are based on centroids of features. In all instances presented here and so far in our own in-house experience, this is sufficient to return any known bridges. Returned bridge 'hits' will have matching SMARTS patterns at their termini separated by D Å. This is illustrated in the context of the full CONFIRM algorithm in Fig. 6.

Connect fragments

The next stage in the CONFIRM algorithm is the connection of the bridge hits to the pre-defined fragment molecules and is done within Pipeline Pilot. The SMARTS search query used with PharmFinder is also used to define the substructure within the fragment molecules. This substructure is matched to the termini on the bridge hits and connection occurs using PilotScript within Pipeline Pilot by overlaying the bridge termini to the anchors on the fragments (Fig. 6). Special care is taken to account for symmetry in the bridge hits and duplicate molecules are removed. A single three dimensional conformer of each connected molecule is generated.

Prepare, dock, and score ligands

The connected molecules are then prepared for docking using LigPrep [57]. The specified chirality, tautomeric form, and ionization state of the input ligand are retained. LigPrep acts as another layer of structure quality filter during this stage of processing as in a limited number of cases, undesirable compounds can still be generated. Representative examples of such failing bridge structures are presented in Fig. 2b. Failing structures may result from unrealistic combinations of chiralities, particularly in fused ring systems which might result from the LigPrep's stereoizer process and then are docked into the target receptor. Molecules processed through LigPrep are then docked into the target receptor using the Glide 4.5 docking software [58–60]

Core restraints, strain energy, and post minimization

The Schrödinger strain_rescore python script [61] was used to compute strain energies of the docked poses as a post-docking step. This script performs a conformational expansion to determine the conformation of the free ligand, followed by computation of the internal energy of the bound and free ligand using the OPLS_2005 force field [62]. The resultant 'strain energy' is the difference between those bound and free energies.

The use of post docking minimization has also been investigated. Within Glide this is only available when

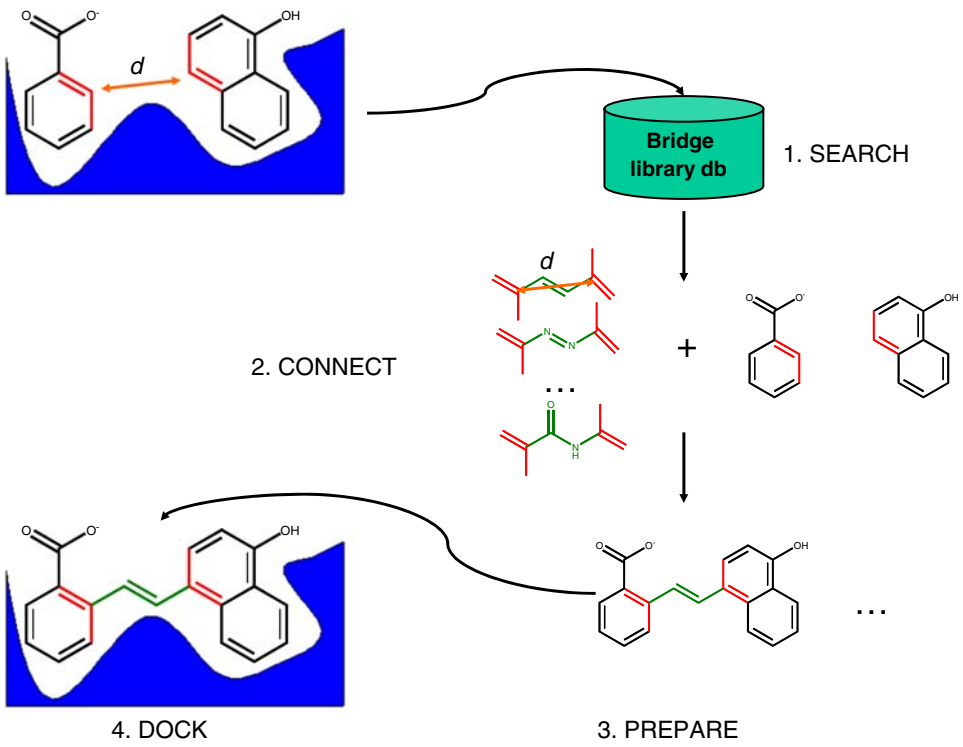


Fig. 6 Schematic detailing the different steps of the CONFIRM algorithm: Search, Connect, Prepare, and Dock

docking in the SP or high-throughput virtual screening (HTVS) modes. Bond lengths, angles, and torsions, are optimized using upto 100 steps of conjugate gradient minimization and the resulting pose is then rescored.

Protein preparation

Three targets were considered: Streptavidin (1SRJ) [63], Sialidase (1A4Q) [64], and Human retinoic acid receptors (1FCZ) [65]. They all have high resolution ($\leq$2.2 Å) protein–ligand X-ray structures available; and represent a variety of ligand chemotypes (Fig. 7). The ligands have two well defined component fragment moieties connected via a bridge. In each structure, the two fragments interact with different binding pockets and are thus suitable test cases to validate the CONFIRM algorithm. Starting from the PDB file, each structure was prepared using the Protein Preparation Wizard [66]; for each complex, water molecules were removed, all hydrogens were added, and His, Asn, and Gln side chain orientations were adjusted as appropriate. The resulting complex structures were used to generate Glide grids for docking; default parameters were used with no constraints, unless otherwise specified.

To validate the receptor preparation stage, for each complex the X-ray ligand was re-docked using both the SP and XP scoring functions, in each case the single top-scoring pose was retained. The heavy atom root-mean-squared deviation (RMSD)[1] from the X-ray bound position of the ligand was computed (Table 2). Excellent agreement was observed between the docked pose and the bioactive conformer (within 1.2 Å of the crystallographic data in all cases), indicating that Glide is a competent choice of docking software for the systems under consideration.

CONFIRM as a de novo design tool

Given that each of the targets are competent to re-dock the known ligand (Table 2), when the bridge matching the ligand structure is present in the library, CONFIRM should be able to re-generate the ligand and dock it correctly. When using CONFIRM as a de novo design tool a combination of the score for the re-constructed ligand and the RMSDs of the two fragments after docking can be used to prioritize new compounds for synthesis. A ligand that is proposed would be rated based on the binding mode of the underlying fragment molecules and their overall score/rank compared to the known ligand or ligands. For de novo design, novel ligands with predicted binding affinity close to, or better than, existing ligands and with fragment position RMSDs less than $\sim$1 Å could be selected for synthetic consideration.

[1] We have used an OEChem utility for computing the RMSD.

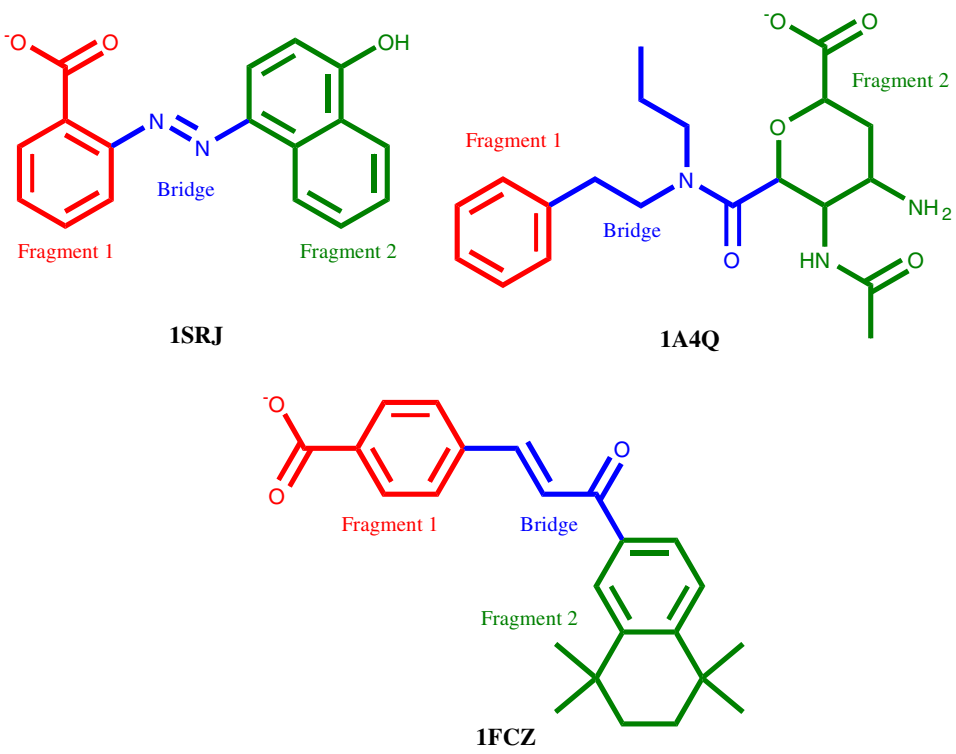


Fig. 7 Native ligands from the complex structures used in the present work. Ligands have been deconstructed into Fragments, along with a bridge

Table 2 Glide SP/XP re-docking of the X-ray bound conformation of the ligand

PDB Ascension code	Resolution (Å)	RMSD (Å)	
		SP	XP
1SRJ[a]	1.80	1.19	0.95
1A4Q[b]	1.90	0.27	0.29
1FCZ[c]	1.38	0.30	0.43

[a] Reference [63]

[b] Reference [64]

[c] Reference [65]

Results and discussion

Streptavidin complex 1SRJ [63]

The naphthyl-HABA ligand, 2-((4′-hydroxynaphthyl)-azo) benzoic acid, binds in the biotin-binding site of streptavidin. The HABA ligand contains a diazene bridge, (Fig. 7) which is included in Lib3.

The connecting through-space distance between sp^2 carbons of the bridge is measured in the X-ray complex structure as 3.70 Å. As a proof of concept of the CONFIRM methodology the smallest library of bridges (Lib3) was searched with a distance query of 4 ± 1.4 Å; this resulted in 29 hits, one of which was the X-ray ligand bridge. Those 29 hits were then automatically connected to the two fragments, benzoic acid and napthalen-1-ol, resulting in 36 ligands that were then prepared for docking. Given the symmetric nature of the two sp^2 carbon attachments, theoretically 58 ligands could result from the connection step. However, 22 are duplicates since the bridges themselves are symmetric and nine additional ligands do not pass the LigPrep stage and are thus removed. The remaining 27 ligands are docked into the crystal structure using Glide. Both the SP and XP scoring schemes have been used and the RMSD of each of the fragments versus the docking score is shown in Figs. 8 and 9, respectively. However, the napthalen-1-ol fragment is generally not positioned as well. Of the 27 docked ligands, the re-constructed X-ray ligand ranks 14th with an overall RMSD of 1.17 Å, and 7th with an RMSD of 0.71 Å with the SP and XP scoring functions, respectively.

In this initial test of CONFIRM the required bridge is found in the smallest bridge library, Lib3, and the known ligand binding mode is reproduced.

Sialidase complex 1A4Q [67]

Zanamivir was the first potent inhibitor of the influenza virus sialidase to be discovered [68]. Amide group substitutions resulted in a series of carboxamide analogues whose properties have been determined both biologically and structurally [64, 69]. One of these, 5-(Acetylamino)-4-

amino-6-(phenethylpropylcarbamoyl)-5,6-dihydro-4H-pyran-2-carboxylic acid (Fig. 7) is described in the 1A4Q complex.

The bridging group of the carboxamide ligand has 7 rotatable bonds and so, it cannot be found in any of the prepared bridge libraries. The unsubstituted N-propylacetamide analog exists in Lib4.

The most similar bridge within Lib3 (similarity 0.79) is just, one carbon shorter (N- ethylacetamide).

The measured through-space sp^2 to sp^3 carbon distance in the bound ligand is 5.91 Å; thus a PharmFinder search query distance of 6 ± 1.4 Å was used. This search criterion was used against all four of the bridge libraries and the results are presented in Table 3. The presence or absence of the 'most similar' bridges propylacetamide or ethylacetamide is noted, along with the number of hits found, the number of unique molecules, the number that fail the LigPrep stage, and the final number submitted to Glide for docking. Only the first occurrence of each unique ligand (as defined by canonical SMILES) is docked. As seen in the previous section, Glide is reasonably insensitive to the input docking conformation. The main observation from Table 3 is that the 4 rotatable bond bridges, N-propylacetamide bridge, is only recovered when the library is conformationally expanded; this highlights the need for conformational sampling when looking at more flexible bridges.

A more detailed analysis of the results presented in Table 3, reveals how preparation of the libraries contributes to the hits found during the PharmFinder step. Of the 143 unique molecules found from searching Lib3E, 15 (10%) are due to the conformational expansion. Similarly, of the 274 unique molecules found from searching Lib4, 146 (53%), have >3 rotatable bonds. Finally, of the 370 unique molecules obtained when searching Lib4E an additional 81 (22%) come from the OMEGA expansion, but have >3 rotors.

The search was repeated whereby Lib4 was seeded with the native bridge from the X-ray ligand. Using the query

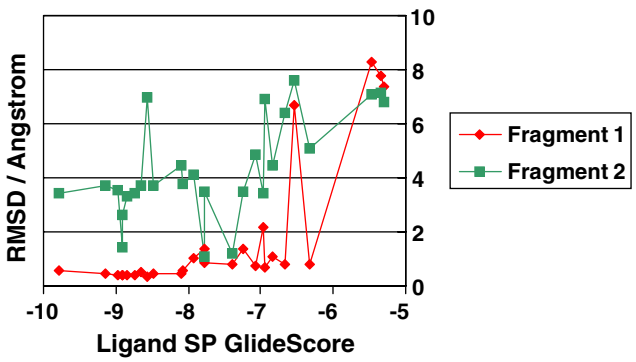


Fig. 8 1SRJ SP GlideScore versus RMSD of Fragments 1 and 2. The benzoic acid fragment is consistently docked well, making multiple interactions with the receptor through the carboxylate group

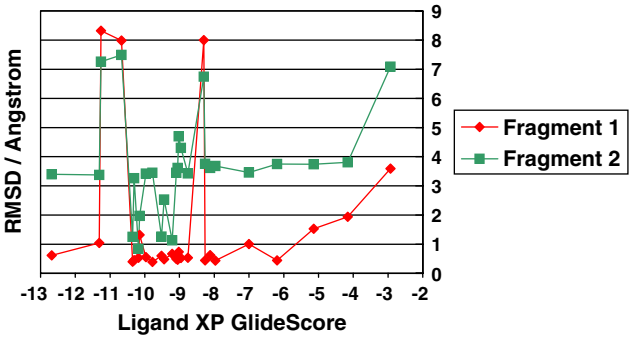
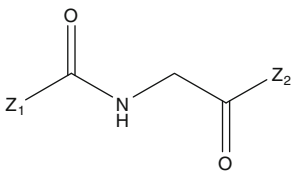


Fig. 9 1SRJ XP GlideScore versus RMSD of Fragments 1 and 2. The benzoic acid fragment is consistently docked well, making multiple interactions with the receptor through the carboxylate group

6 ± 1.4 Å and searching against this 'seeded Lib4', the additional hit, the seeded bridge, is recovered (Fig. 10).

In Fig. 11, for each of the bridge libraries searched against, we have computed the percentage of docked ligands with an RMSD of <2 Å for both fragments. A concomitant increase in this percentage measure is observed as the bridge library is conformationally expanded and thereby bridge flexibility is included. At least for this example, the XP scoring function seems to generate more poses where both fragments are optimally placed. While the 'fragment RMSD <2 Å' metric is useful, we found that adding a cutoff based on docking score employed simultaneously, improved the quality of the selection of bridges that can be suggested to the chemists. If compared to the docking score of the X-ray ligand, any new ligand with better scoring score should in theory have a better potency than the X-ray ligand. In our case, this happens if the new bridge either realized new interaction(s) with the receptor or improves the rigidity of the ligand. This is the case with an N-(2-oxo-propyl)-acetamide bridge we found while screening the Lib4E library.



Human retinoic acid receptors

CONFIRM is also applied to a number of retinoid agonists which are efficacious against retinoic acid receptors. Initially CONFIRM was used to regenerate the native ligand of 1FCZ [65] (Fig. 7), and then to investigate a series of ligands each of which contains the same fragments as in 1FCZ, but with different bridges for which experimental binding affinities (K_Ds) are available. All of these ligands are agonists which are preferentially selective for the γ isotype of the human retinoic acid receptor. This series of

Table 3 Search results[a] for 1A4Q against the four bridge libraries

Library	No. of hits	No. of unique hits	No. of ligands removed by LigPrep	No. of docked ligands	Ethyl amide bridge present?	Propyl amide bridge present?
Lib3	128	128	3[b]	125	Yes	NA
Lib3E	544	143	3[b]	140	Yes	NA
Lib4	274	274	4[c]	270	Yes	No
Lib4E	2314	370	4[c]	366	Yes	Yes

[a] Using a query of 6 ± 1.4 Å

[b] The same three structures are removed

[c] The same four structures are removed

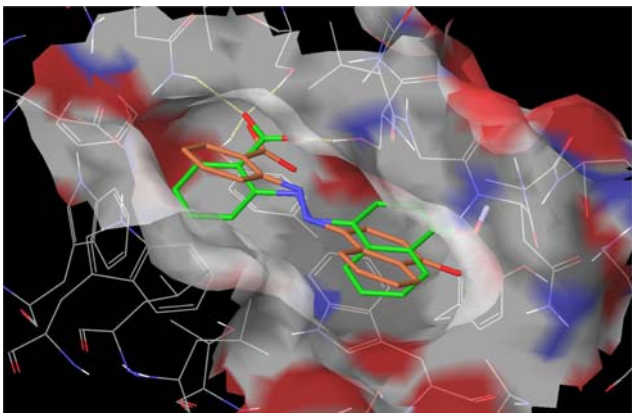


Fig. 10 The bound conformation of the X-ray ligand (green) shown with the SP pose of the re-constructed X-ray ligand (brown)

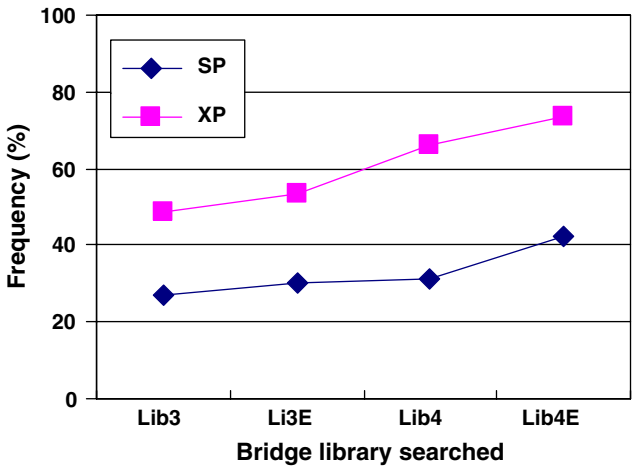


Fig. 11 The frequency of hits (as a percentage) which have an RMSD of Fragment 1 and 2 less than 2 Angstrom as a function of bridge library searched

ligands has been investigated both in terms of the ability of CONFIRM to regenerate the X-ray ligand and to evaluate different scoring function available through Glide to define the best scoring option to use in CONFIRM, to rank order ligand correctly.

1FCZ [65]

The human retinoic acid receptors (hRAR) belongs to the family of nuclear hormone receptors and are involved in the regulation of cell growth, differentiation and apoptosis. The RAR family is composed of three isotypes α, β, and γ which are distinct pharmacological targets for the design of retinoids for treatment of various skin diseases and cancers. A sequence alignment of the three RAR isotypes shows that all but three residues in the ligand-binding site are conserved [70–72], and the importance of these residues for isotype selectivity has been previously demonstrated [73]. In particular, selectivity against RARγ can be achieved through interactions with Met 272 [65, 71, 74].

The 1FCZ ligand shown in Fig. 7, BMS181156, is a RAR pan-agonist being equally potent against the α, β, and γ isotypes (K_D values of 0.4, 3.5, and 0.6 nM respectively [65]). The through-space sp^2 carbon to sp^2 carbon distance in BMS181156 is measured to be 4.55 Å in the X-ray structure, and a subsequent PharmFinder search using a 5 ± 1.4 Å query with Lib3 results in 148 hits, one of which is the X-ray ligand pent-3-en-2-one bridge. Subsequent connection of the bridges to the fragments results in 262 complete molecules, of which 217 are successfully processed through LigPrep. These are docked into the 1FCZ hRAR receptor conformation using both the SP and XP scoring functions, and the results for the re-constructed X-ray ligand are shown in Table 4. The SP function places the ligand containing FFDPZHA within the top 5% of the molecules scored while XP places it within the top 50%; for both scoring functions the RMSD of the associated resulting poses is excellent (<0.5 Å).

1FCX, 1FCY and 1FD0

The ligands from the hRARγ complexes 1FCX, 1FCY, and 1FD0 are docked into the 1FCZ protein structure. These ligands are detailed in Fig. 12 and all have the common

Table 4 Glide SP/XP docking for re-constructed X-ray ligands

PDB Ascension code	Bridge present?	SP		XP	
		RMSD (Å)	Rank	RMSD (Å)	Rank
1SRJ	Yes	1.17	14/27	0.71	7/27
1A4Q	No	NA		NA	
1FCZ	Yes	0.47	10/205[a]	0.49	107/215[b]

[a] 12 poses rejected by energy filter

[b] 2 poses rejected by energy filter

fragments: Fragment 1—5,6,7,8-tetrahydro-5,5,8,8-tetramethyl-2-naphthalenyl (TTN), and Fragment 2—2-napthoic acid, but differ in the connecting bridges.

In the 1FCX ligand the bridge consists of a hydroxyl moiety attached at a chiral center; while the 1FCX ligand is a RARγ selective agonist, the *R*-enantiomer exhibits ten-fold higher selectivity versus the *S*-enantiomer. At the bridging position the 1FCY ligand has a keto group and the 1FD0 ligand bridge is an oxime. In the case of the 1FD0 ligand two possible orientations of the oxime group exist (corresponding to the E and Z isomers). Both isomeric states are found in similar proportions in the X-ray structure and this suggests similar affinities; although individual measurements have not been made, the mixture has a K_D of 3-5nM. In the 1FD0 crystal structure, the Met 272 residue moves to form a critical interaction with the hydroxyl group moiety of the oxime bridge (Fig. 13). The synergy of classical and CH···O hydrogen bonds are used to hypothesize the dual potency of the 1FD0 isomers [75]. A similar motion of the Met 272 is also observed in the crystal structure of 1FCX (where the oxime group is replaced by a hydroxyl group); although no additional CH···O hydrogen bonds are observed. Figure 13 shows an overlay of the

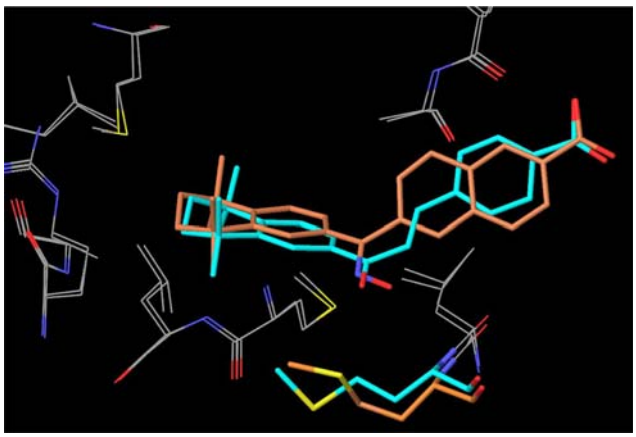
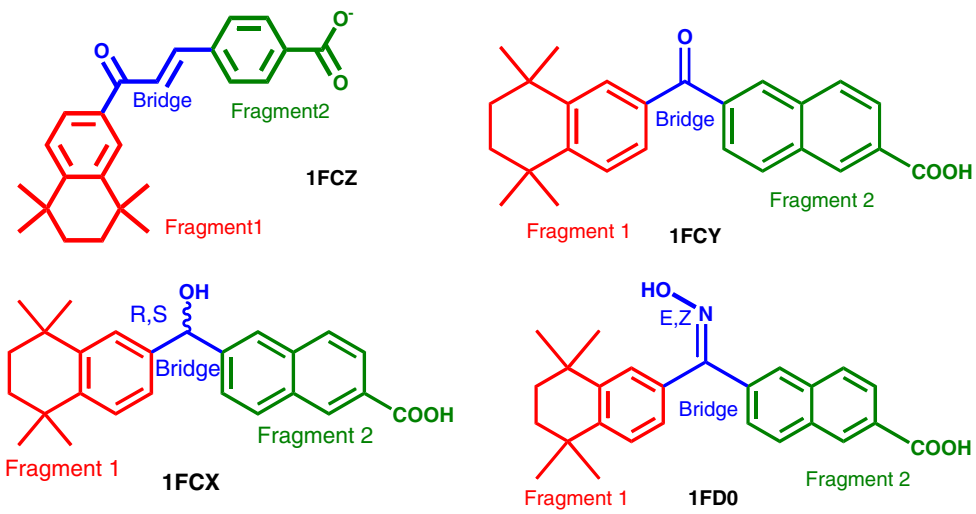


Fig. 13 Overlay of 1FCZ (cyan) and 1FD0 (brown). Residue Met 272 is highlighted to illustrate the interaction between the sulfur atom and the keto (1FCZ) and hydroxyl (1FD0) bridging groups. In this latter example the E isomer is shown

hRARγ proteins from the 1FCZ and 1FD0 complexes. One would hope that a docking experiment of these ligands against the protein conformation of 1FCZ, and against the protein conformation of 1FD0, would result in different rank ordering of ligands assuming that the scoring function was able to correctly represent the physical interactions; although this is a particularly challenging example given the highly non-classical nature of the CH···O interaction.

An analysis of Lib3 reveals the presence of only the 1FCY bridge so the missing 1FCX and 1FD0 ligand bridges are seeded into this library. The measured through-space sp^2 carbon to sp^2 carbon distances are 2.59 Å (1FCX), 2.58 Å (1FCY), and 2.57 Å (1FD0), respectively, and a PharmFinder search with the query 3 ± 1.4 Å against this seeded Lib3 library results in 24 hits (all of the seeded bridges are returned). Subsequent connection of the bridges to the fragment molecules results in 26 complete molecules, 23 of these were processed successfully through

Fig. 12 Native ligands to complexes 1FCZ, 1FCX, 1FCY [65], and 1FD0 [75]



LigPrep (multiple stereoisomers are generated to ensure that both the Z isomer of 1FD0 and the *S* enantiomer of 1FCX are present). These 23 molecules are then docked into the 1FCZ structure using Glide and scored using both the SP and XP scoring functions. Within the SP scoring, of the 23 molecules scored, three of them fail one (or more) components of the Glide Score evaluation and are thus automatically discarded by Glide. These are not considered for the remainder of this analysis.

In addition to the SP and XP scoring functions an MM-GBSA [76] and MM-PBSA [77] schemes were used to re-score docked SP poses. For each of these scoring schemes the RMSD of the docked ligand (with respect to the bound position), and the predicted rank of the compound in the list of 20 was determined. The binding affinity data extant in the literature ranks the ligands accordingly: BMS181156 (1FCZ) > 1FCY ~ 1FD0 > 1FCX (for 1FCX individual affinities have been measured for both the *S* and *R* enantiomers, however the racemic mixture has an affinity of 64 nM and is RARγ selective). Except XP, all of the methods rank the BMS181156 ligand 1st out of 20, and all poses (for all methods) have an RMSD of <0.5 Å. The incorrect ranking of the 1FD0 and 1FCX ligands is to be expected given the substantial movement of Met 272 (Table 5).

The CONFIRM procedure was repeated, using the 1FD0 receptor and the same 20 ligands with docking and scoring as described above. Results summarized in Table 6 show poor ranking, reflecting the underlying importance of the interaction between the bridging moiety in each of these ligands and the sulfur on Met 272, which has important implications for designing RARγ selective inhibitors. It also illustrates the importance of the conformation of the receptor when performing detailed lead optimization, especially when it is noted that the 1FD0 (E and Z) ligands, while being equipotent with 1FCY, were predicted to be poor binders when docking against the 1FCZ conformation of hRARγ; the movement of the sulfur atom is critical both from a selectivity and an affinity perspective. It is also clear given the large difference between the 1FD0 isomer rankings that the CH···O bonding network which has been well hypothesized experimentally, is not well captured from a scoring perspective.

Conclusions

CONFIRM, a novel and fully automated method for the connection of fragment molecules has been introduced. This algorithm involves searching a pre-prepared library of bridging fragments with a SMARTS substructure pattern search engine to extract 'hit' molecules matching an input criterion. These hits are then connected to pre-defined fragment molecules (with well resolved positions) and the

Table 5 Rank and scoring information for the hRAR ligands[a]

		K_D/nM	SP		XP		Prime/MM-GBSA		MM-PBSA	
			Rank	RMSD (Å)	Rank	RMSD (Å)	Rank	RMSD (Å)	Rank	RMSD (Å)
BMS181156		0.6	1	0.30	10	0.43	1	0.27	1	0.27
1FCY		3	2	0.66	12	0.53	2	0.48	5	0.62
1FCX	S	816	7	0.58	9	0.61	14	0.51	6	0.57
	R	75	8	0.61	14	0.71	13	0.54	4	0.58
1FD0	E	3–5	18	0.66	3	0.55	17	0.36	16	0.54
	Z		10	0.71	13	0.56	16	0.52	18	0.65

[a] Docked to the 1FCZ receptor

Table 6 Rank and scoring information for the hRAR ligands[a]

		K_D/nM	SP		XP		Prime/MM-GBSA		MM-PBSA	
			Rank	RMSD (Å)	Rank	RMSD (Å)	Rank	RMSD (Å)	Rank	RMSD (Å)
BMS181156		0.6	10	0.36	12	0.26	10	0.28	4	0.19
1FCY		3	4	0.48	14	0.40	2	0.35	8	0.50
1FCX	S	816	7	0.46	16	0.46	14	0.48	3	0.47
	R	75	6	0.57	3	0.67	15	0.53	12	0.46
1FD0	E	3–5	19	0.30	8	0.37	6	0.18	9	0.42
	Z		5	0.59	2	0.61	12	0.49	10	0.45

[a] Docked to the 1FD0 receptor

resulting ligands are then prepared and docked. Analysis of the docked poses has been performed in terms of both the RMSD from the known fragment positions and docking score in the case of known inhibitors.

Several examples from the PDB were used to demonstrate the efficacy of the CONFIRM methodology. These examples all have well defined binding cavities, excellent crystallographic resolution, and native ligands composed of two distinct fragments and a bridging group. The bridge libraries prepared for use with CONFIRM contain bridges found across a diversity of small molecules. It has also been demonstrated that the method used for extracting bridge 'hits' from a database (PharmFinder) is robust and insensitive to the input criterion. The CONFIRM algorithm leverages Glide as a docking engine, and it has been shown that it is capable of both generating poses that reproduce known binding modes and, in certain instances, competency at rank ordering ligands with known binding affinity even using the simplest standard precision scoring function. The main issue we encounter in CONFIRM was its limitation on dealing with flexibility in the receptor as shown in the human Retiniods example. But this draw back is not surprising since it is still one of the main limitations in docking methodology.

Given this validation of CONFIRM this methodology is now being applied to drug discovery programs in-house. It is anticipated that wherever fragment-based structural information is available (either experimentally derived, or through well-validated computational experiments) CONFIRM could be utilized to great effect. The ability to quickly optimize a fragment into a region of chemical space that has the necessary potency and physical property profiles for the medicinal chemistry project is well documented in the literature (see for example reference [21]); a simple tool such as CONFIRM offers an interesting, novel, approach to the problem. Furthermore, its inherently modular design affords project specific modifications that we have already begun to explore in-house. Although, as with any de novo tool, its real utility will only be explored when compounds suggested through its use are synthesized, assayed, and shown experimentally to bind to the target of interest as predicted.

Acknowledgement We would like to thank J. Christian Baber for fruitful discussions during the preparation of this manuscript.

References

1. Hertzberg RP, Pope AJ (2000) Curr Opin Chem Biol 4:445
2. Macarron R (2006) Drug Discov Today 11:277
3. Oprea TI, Matter H (2004) Curr Opin Chem Biol 8:349
4. Shoichet BK (2004) Nature 432:862
5. Joseph-McCarthy D (1999) Pharmacol Therapeut 84:179
6. Joseph-McCarthy D, Baber JC, Feyfant E, Thompson DC, Humblet C (2007) Curr Opin Drug Discov Devel 10:264
7. Jencks WP (1981) Proc Natl Acad Sci USA 78:4046
8. Farmer PS, Ariens EJ (1982) Trends Pharmacol Sci 3:362
9. Abad-Zapatero C, Metz JT (2005) Drug Discov Today 10:464
10. Hopkins AL, Groom CR, Alex A (2004) Drug Discov Today 9:430
11. Hajduk PJ (2006) J Med Chem 49:6972
12. Pellecchia M, Sem DS, Wuthrich K (2002) Nat Rev Drug Discov 1:211
13. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Science 274:1531
14. Blundell TL, Jhoti H, Abell C (2002) Nat Rev Drug Discov 1:45
15. Kuhn P, Wilson K, Patch MG, Stevens RC (2002) Curr Opin Chem Biol 6:704
16. Hajduk PJ, Sheppard G, Nettesheim DG, Olejniczak ET, Shuker SB, Meadows RP, Steinman DH, Carrera GM, Marcotte PA, Severin J, Walter K, Smith H, Gubbins E, Simmer R, Holzman TF, Morgan DW, Davidsen SK, Summers JB, Fesik SW (1997) J Am Chem Soc 119:5818
17. Oprea TI, Davis AM, Teague SJ, Leeson PD (2001) J Chem Inf Comp Sci 41:1308
18. Carr RAE, Congreve M, Murray CW, Rees DC (2005) Drug Discov Today 10:987
19. Rees DC, Congreve M, Murray CW, Carr R (2004) Nat Rev Drug Discov 3:660
20. Schneider G, Fechner U (2005) Nat Rev Drug Discov 4:649
21. Hajduk PJ, Greer J (2007) Nat Rev Drug Discov 6:211
22. Allen KN, Bellamacina CR, Ding XC, Jeffery CJ, Mattos C, Petsko GA, Ringe D (1996) J Phys Chem 100:2605
23. Joseph-McCarthy D, Fedorov AA, Almo SC (1996) Protein Eng 9:773
24. Howard N, Abell C, Blakemore W, Chessari G, Congreve M, Howard S, Jhoti H, Murray CW, Seavers LCA, van Montfort RLM (2006) J Med Chem 49:1346
25. Danziger DJ, Dean PM (1989) Proc R Soc Lond B Biol Sci 236:101
26. Gillet VJ, Myatt G, Zsoldos Z, Johnson AP (1995) Perspect Drug Discov Design 3:34
27. Miranker A, Karplus M (1991) Proteins 11:29
28. Evensen E, Joseph-McCarthy D, Karplus M, Weiss GA, Schreiber SL, Karplus M (2007) J Comput Aided Mol Des 21(7):395
29. Gillet V, Johnson AP, Mata P, Sike S, Williams P (1993) J Comput Aided Mol Des 7:127
30. Mata P, Gillet VJ, Johnson AP, Lampreia J, Myatt GJ, Sike S, Stebbings AL (1995) J Chem Inf Comp Sci 35:479
31. Nishibata Y, Itai A (1991) Tetrahedron 47:8985
32. Bohm HJ (1993) J Mol Recognit 6:131
33. Bohacek RS, McMartin C (1994) J Am Chem Soc 116:5560
34. Wang R, Gao Y, Lai L (2000) J Mol Model 6:498
35. Todorov NP, Dean PM (1997) J Comput Aided Mol Des 11:175
36. Todorov NP, Dean PM (1998) J Comput Aided Mol Des 12:335
37. Ishchenko AV, Shakhnovich EI (2002) J Med Chem 45:2770
38. Murray CW, Verdonk ML (2002) J Comput Aided Mol Des 16:741
39. Miranker A, Karplus M (1995) Proteins 23:472
40. Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, Waszkowycz B, Westhead DR (1995) J Comput Aided Mol Des 9:13
41. Ho CMW, Marshall GR (1993) J Comput Aided Mol Des 7:623
42. Lauri G, Bartlett PA (1994) J Comput Aided Mol Des 8:51
43. Yang YL, Nesterenko DV, Trump RP, Yamaguchi K, Bartlett PA, Drueckhammer DG (2005) J Chem Inf Model 45:1820
44. Sefler AM, Kozlowski MC, Guo T, Bartlett PA (1997) J Org Chem 62:93
45. Smith WW, Bartlett PA (1998) J Am Chem Soc 120:4622

46. Weiss GA, Collins EJ, Garboczi DN, Wiley DC, Schreiber SL (1995) Chem Biol 2:401
47. Artis DR, Brotherton-Pleiss C, Pease JH, Lin CJ, Ferla SW, Newman SR, Bhakta S, Ostrelich H, Jarnagin K (2000) Bioorg Med Chem Lett 10:2421
48. Takano Y, Koizumi M, Takarada R, Kamimura MT, Czerminski R, Koike T (2003) J Mol Graph Model 22:105
49. Nilakantan R, Nunn DS, Greenblatt L, Walker G, Haraki K, Mobilio D (2006) J Chem Inf Model 46:1069
50. SciTegic Pipeline Pilot, version 6.1, Accelrys, Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121 USA
51. Nilakantan R, Immermann F, Haraki K (2002) Comb Chem High Throughput Screen 5:105
52. Nilakantan R, Nunn DS (2003) Drug Discov Today 8:668
53. Baber JC, Feher M (2004) Mini-Rev Med Chem 4:681
54. OMEGA2.0 (2006) OpenEye. Santa Fe, New Mexico
55. Chemical Computing Group Inc. (2005) Montreal, Canada
56. Daylight Chemical Information Systems, Inc. (2006) Aliso Viejo, California
57. LigPrep 2.1 (2005) Schrödinger. LLC, New York, NY
58. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) J Med Chem, ASAP
59. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) J Med Chem, ASAP
60. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) J Med Chem 49:6177
61. Revision 1.1.2.2 (2006) Schrödinger. LLC, New York, NY
62. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA (2002) J Comput Chem 23:1515
63. Weber PC, Pantoliano MW, Simons DM, Salemme FR (1994) J Am Chem Soc 116:2717
64. Taylor NR, Cleasby A, Singh O, Skarzynski T, Wonacott AJ, Smith PW, Sollis SL, Howes PD, Cherry PC, Bethell R, Colman P, Varghese J (1998) J Med Chem 41:798
65. Klaholz BP, Mitschler A, Moras D (2000) J Mol Biol 302:155
66. Revision 1.1.2.6 (2006) Schrödinger. LLC, New York, NY
67. von Itzstein M, Wu WY, Kok GB, Pegg MS, Pegg MS, Dyason JC, Jin B, von Phan T, Smythe ML, White HF, Oliver SW et al (1993) Nature 363:418
68. Weber PC, Wendoloski JJ, Pantoliano MW, Salemme FR (1992) J Am Chem Soc 114:3197
69. Smith PW, Sollis SL, Howes PD, Cherry PC, Starkey ID, Cobley KN, Weston H, Scicinski J, Merritt A, Whittington A, Wyatt P, Taylor N, Green D, Bethell R, Madar S, Fenton RJ, Morley PJ, Pateman T, Beresford A (1998) J Med Chem 41:787
70. Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, Moras D (1995) Nature 378:681
71. Klaholz BP, Renaud JP, Mitschler A, Zusi C, Chambon P, Gronemeyer H, Moras D (1998) Nat Struct Biol 5:199
72. Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, Gronemeyer H (1996) Nat Struct Biol 3:87
73. Gehin M, Vivat V, Wurtz JM, Losson R, Chambon P, Moras D, Gronemeyer H (1999) Chem Biol 6:519
74. Klaholz BP, Mitschler A, Belema M, Zusi C, Moras D (2000) Proc Natl Acad Sci USA 97:6322
75. Klaholz BP, Moras D (2002) Structure 10:1197
76. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) J Med Chem 49:5912
77. Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP (2006) J Chem Inf Model 46:243