THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

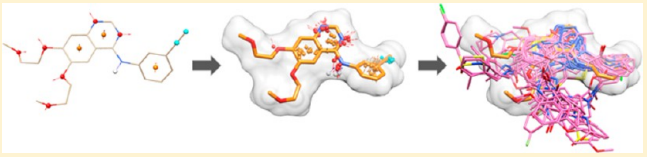
pubs.acs.org/JPCB

Pharmacophore-Based Similarity Scoring for DOCK

Lingling Jiang[†] and Robert C. Rizzo*[,†,‡,§]

[†]Department of Applied Mathematics & Statistics, [‡]Institute of Chemical Biology & Drug Discovery, [§]Laufer Center for Physical & Quantitative Biology, Stony Brook University, Stony Brook, New York 11794-3600, United States

ABSTRACT: Pharmacophore modeling incorporates geometric and chemical features of known inhibitors and/or targeted binding sites to rationally identify and design new drug leads. In this study, we have encoded a three-dimensional pharmacophore matching similarity (FMS) scoring function into the structure-based design program DOCK. Validation and characterization of the method are presented through pose reproduction, crossdocking, and enrichment studies. When used alone, FMS scoring dramatically improves pose reproduction success to 93.5% (~20% increase) and reduces sampling failures to 3.7% (~6% drop) compared to the standard energy score (SGE) across 1043 protein−ligand complexes. The combined FMS +SGE function further improves success to 98.3%. Crossdocking experiments using FMS and FMS+SGE scoring, for six diverse protein families, similarly showed improvements in success, provided proper pharmacophore references are employed. For enrichment, incorporating pharmacophores during sampling and scoring, in most cases, also yield improved outcomes when docking and rank-ordering libraries of known actives and decoys to 15 systems. Retrospective analyses of virtual screenings to three clinical drug targets (EGFR, IGF-1R, and HIVgp41) using X-ray structures of known inhibitors as pharmacophore references are also reported, including a customized FMS scoring protocol to bias on selected regions in the reference. Overall, the results and fundamental insights gained from this study should benefit the docking community in general, particularly researchers using the new FMS method to guide computational drug discovery with DOCK.



1. INTRODUCTION

Many docking and virtual screening programs, such as DOCK,[1,2] employ intermolecular interaction energy functions that contain nonbonded van der Waals and electrostatic terms to rank-order (i.e., score) small molecule binding geometries (poses) generated in the context of a defined protein binding site. Other physically reasonable scoring terms such as intermolecular hydrogen-bonding, ligand desolvation, numbers of ligand rotatable bonds, and buried surface area, among others, have also been explored.[3] In all cases, the objective is to enrich for ligands with good geometric and chemical compatibility with the target so that promising druglike leads can be identified.[4−6] Recently, Balius et al.[7,8] reported a new DOCK scoring method termed footprint similarity score which can be used to identify compounds that match a specific molecular interaction energy pattern (i.e., footprint) based on a known reference ligand. Encouraged by the recent successes[9,10] from our laboratory, in which "footprints" were used to identify promising lead compounds, we have developed an analogous similarity-based scoring method for DOCK that employs "pharmacophores". Both methods yield enhanced docking outcomes but do so in an orthogonal sense (energy vs geometry).

Historically, the concept of a pharmacophore is generally attributed to Ehrilich[11,12] and has evolved to include the three-dimensional spatial arrangements of key chemical features essential for compound affinity leading to a biological effect.[13,14] A thorough summary of the development of pharmacophores and early works on modeling can be found in a recent publication by Güner et al.[13] Reviews by Leach et al.,[15] Yang,[16] and Sanders et al.[17] also discuss technological advances and challenges of using different pharmacophore methods in modern drug discovery. In practice, pharmacophore features can be derived from known active ligand(s), a defined binding site geometry, or a combination of both. Importantly, the abundance of atomic-resolution structures publically available in the protein data bank (PDB)[18] can be used to derive pharmacophore models for compounds with verified experimental activity to help guide structure-based drug design. A partial list of programs that incorporate pharmacophore modeling includes CATALYST,[19] GASP,[20] LigandScout,[21] PHASE,[22] GALAHAD,[23] PhDOCK,[24,25] and MOE,[26] among others. While such prior efforts are important tools and represent different approaches for modeling, the goal of the present work is to provide a pharmacophore method that can leverage DOCK's powerful anchor-and-grow sampling strategy while taking advantage of different combinations of scoring functions.

The new DOCK pharmacophore scoring protocol termed Pharmacophore Matching Similarity (FMS) encodes useful chemical features, including hydrogen bond acceptors/donors, hydrophobic groups, positively/negatively charged groups, and aromatic/nonaromatic rings. Initial pharmacophore types are generated based on atom type and chemical environment, defined by neighboring atoms in the same ligand molecule, and

Special Issue: William L. Jorgensen Festschrift

Received: July 1, 2014
Revised: September 12, 2014
Published: September 17, 2014

   dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

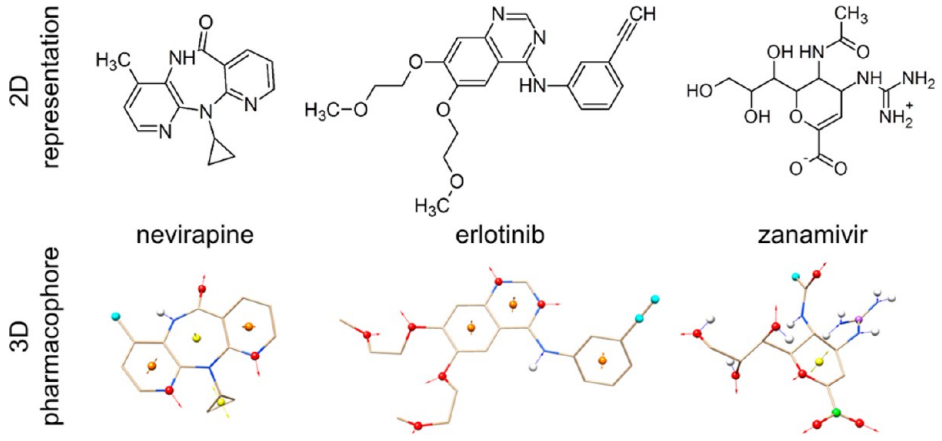


Figure 1. Two-dimensional (2D) representations for three approved drugs (top) and corresponding DOCK pharmacophore (ph4) models (bottom). Features include: (i) hydrogen bond acceptor in red, (ii) hydrogen bond donor in blue, (iii) hydrophobic atom/group in cyan, (iv) aromatic ring center and direction in orange, (v) nonaromatic ring center and direction in yellow, (vi) negatively charged group center in green, and (vii) positively charged group center in magenta. Structures of nevirapine, erlotinib, and zanamivir from PDB codes 1VRT, 1M17, and 1A4G, respectively.

are processed to create a pharmacophore feature set (ph4 model) with coordinates and directionality as shown in Figure 1 for three representative druglike compounds. Importantly, the amount of overlap (termed FMS score) between a user reference ligand pharmacophore and candidate pharmacophores derived from docked compounds can be computed on-the-fly during docking (or rescoring) without the need for a separate preprocessing step. This enables large virtual screening libraries to be sorted (i.e., rank-ordered) with the function in an efficient manner.

Specific validation tests used in this work to evaluate the new scoring protocol include pose reproduction, crossdocking, and enrichment. All FMS results are compared relative to using the standard DOCK single grid energy (SGE) approach, as well as a combined scoring function (FMS+SGE) consisting of both terms. In pose reproduction, crystallographic ligand positions are used as a reference to test if a given method is capable of reproducing nativelike poses (within 2 Å of the X-ray pose) using the large SB2012 validation database (update of SB2010)[27] developed in our laboratory. In crossdocking, select protein families from SB2012 (based on high sequence homology), are employed to evaluate docking accuracy across an $N \times N$ matrix when all ligands from a family are docked to each individual receptor. In enrichment, active ligands and accompanying decoy compounds taken from the DUD-E[28] database are docked to 15 different targets to assess the ability of the new scoring schemes to correctly rank-order active ligands earlier than decoys. Finally, retrospective analyses of three virtual screens to targets of pharmaceutical interest (EGFR, IGF-1R, and HIVgp41) are shown in which FMS-based scoring (FMS and FMS+SGE) was used as a data-mining tool to identify compounds with high pharmacophore overlap to small molecules or peptide ligand side chains. Overall, the results of this comprehensive study suggest the new method will be a useful addition to the growing number of scoring and sampling methods available in DOCK.

2. THEORETICAL METHODS

2.1. Pharmacophore Definitions. Pharmacophore modeling in this study uses a two-step protocol involving: (1) assignment of a pharmacophore type definition to each ligand atom, followed by (2) construction of pharmacophore points with pharmacophore labels based on the type definitions. Inspired by chemical matching code previously developed for DOCK,[24,29] we employ a type definition model based on SYBYL[30] atom types and environment (neighboring atoms). The finite list of pharmacophore type definitions is stored in the *ph4.defn* parameter file (Table 1) and can be customized to include other pharmacophore types. For clarity, it is important to emphasize there is a distinction between pharmacophore type definitions (for the individually typed atoms) and the pharmacophore label definitions (for the final constructed pharmacophore points) derived from the pharmacophore types. In the atom environment definition list in Table 1, parentheses ( ) specify "atoms that must be bonded to the parent atom", while square brackets [ ] specify "atoms that must not be bonded to the parent atom".[31] The integer in the definition represents the number of atoms associated in the rule. For example, the syntax "N.pl3 (2 *) [H]" specifies a trigonal planar nitrogen connected to at least two other atoms and not bound to any hydrogen atoms. For this work, eight pharmacophore types are assigned to individual atoms as outlined in Table 1: (1) null or no assignment, (2) hydrophobic, (3) hydrogen bond donor, (4) hydrogen bond acceptor, (5) aromatic ring member, (6) hydrogen bond acceptor in an aromatic ring, (7) negatively charged species, and (8) positively charged species. The resulting atom set is postprocessed to generate pharmacophore points with coordinates that specify the position of the pharmacophore point center and vectors indicating the direction of potential interactions.

Aromatic and nonaromatic rings are identified by checking for closed loops formed by connected atoms. The coordinate of the ring center, averaged over all ring member coordinates, is computed and saved as the pharmacophore point (Figure 2). The average normal vector of the plane defined by adjacent ring center-to-vertex vectors (Figure 2, dashed blue lines) is calculated and saved as the direction vector of the pharmacophore point. If the individual normal vectors (Figure 2, solid blue lines) of the ring are all within an angle cutoff θ_c to the average normal vector (Figure 2a, solid red lines) then the pharmacophore point is marked as an aromatic ring (Figure 2a). Otherwise it is labeled as nonaromatic (Figure 2b). In practice, θ_i is measured by directly computing the inner product

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

Table 1. Pharmacophore Type Definitions in DOCK

type name[a]	environment definition[b]
(1) null	*
(2) hydrophobic	C. [O.] [N.] [S.] [F] [P] (*)
	C. (N.pl3 (2 C.)) (*)
	N.pl3 (3 C.)
(3) donor	H (O.)
	H (N.)
	H (S.)
	H (F)
(4) acceptor	O. (*)
	N.1 (1 *)
	N.2 [3 *]
	N.3 (3 *)
	N.pl3 (2 *) [H]
	S.2 [O.] [N.]
	S.3 (2 *)
	F (*)
	Cl (*)
(5) aromatic	C.ar
	N.ar
(6) aroAcc	N.ar [H] [3 *] (*)
(7) negative	C. (2 O.co2)
	C.2 (O.2) (O.3 [*])
	P. (4 O.) (O.3 [*])
	S. (3 O.) (O.3 [*])
	S. (4 O.) (O.3 [*])
	F [*]
	Cl [*]
(8) positive	C.cat (*)
	N.4 (*)
	N.3 (4 *)
	N.2 (3 *)
	Zn [*]
	Mg [*]
	Ca [*]
	Mn [*]
	K [*]
	Fe [*]

[a]Types defined in DOCK *ph4.defn* parameter file. [b]Environments based on SYBYL atom types and atom connectivities.

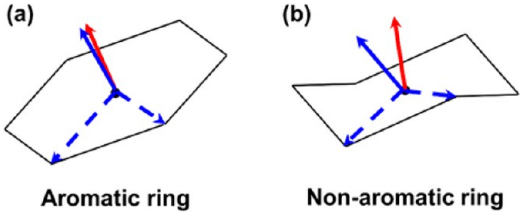
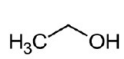
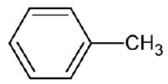
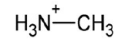
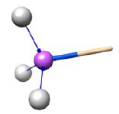
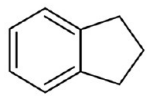
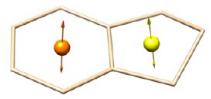
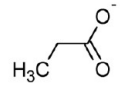
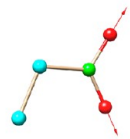
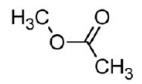
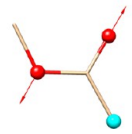


Figure 2. Pharmacophore feature assignment for rings: (a) aromatic (close to planar) and (b) nonaromatic (not planar). Ring center-to-vertex vectors shown as dashed blue lines, individual normal vector shown as solid blue lines, and averaged normal vectors shown in solid red lines. The angle between the blue and the red vectors are compared to a threshold to determine the planarity of the ring.

of two vectors (x_i) which is converted to degrees by the inverse function of cosine as $\arccos(x_i) = \theta_i$. On the basis of examining crystallographic ligand coordinates containing aromatic and nonaromatics rings, we use as a cutoff criteria $\arccos(0.99) \approx 8.11$ degrees to determine if a ring is planar.

Atoms with hydrophobic and positive/negative pharmacophore type definitions are saved individually as pharmacophore points inheriting the same type as their pharmacophore labels. For these cases, default direction vectors (which do not affect the score) are assigned to facilitate a common data structure. For the hydrogen bond acceptor, the coordinate of the polar atom is saved as the pharmacophore point. The average of vectors pointing from all neighbor atoms to the acceptor atom is saved as the direction vector, indicating the potential position of the coupling hydrogen bond donor, as indicated by red arrows in Table 2, which shows example pharmacophores

Table 2. Examples of Pharmacophore Features Derived from Small Molecules

Name	Label[a]	2D[b]	3D[c]
(a) ethanol	PHO HBD HBA		
(b) toluene	ARO PHO		
(c) methyl-amine	POS HBD		
(d) indane	ARO RNG		
(e) propanoic acid	PHO NEG HBA		
(f) methyl acetate	PHO HBA		

[a]PHO in cyan, HBA (vertex and vector) in red, HBD vector in blue, hydrogen vertex in gray, ARO (vertex and vector) in orange, RNG (vertex and vector) in yellow, POS in magenta, and NEG in green. Direction vectors are shown in arrows generated using Chimera[32] bild files. [b]2D Pictures generated with ChemSketch.[33] [c]Three-dimensional (3D) molecules and pharmacophore visualization generated with Chimera

derived for several small organic molecules. The hydrogen bond donor uses the coordinate of the hydrogen atom as that of the pharmacophore point. Similarly, the vector pointing from the donor hydrogen to the connecting polar atom is saved as a normalized direction vector, indicated as blue arrows in Table 2. The combined set of all pharmacophore points is called the molecular pharmacophore (ph4) model, which may include hydrophobic (PHO), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), aromatic ring (ARO), nonaromatic ring (RNG), positively charged (POS), and negatively charged (NEG) features (see Table 2).

To gauge how many pharmacophore features are present in typically sized compounds, Figure 3 plots histograms derived

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

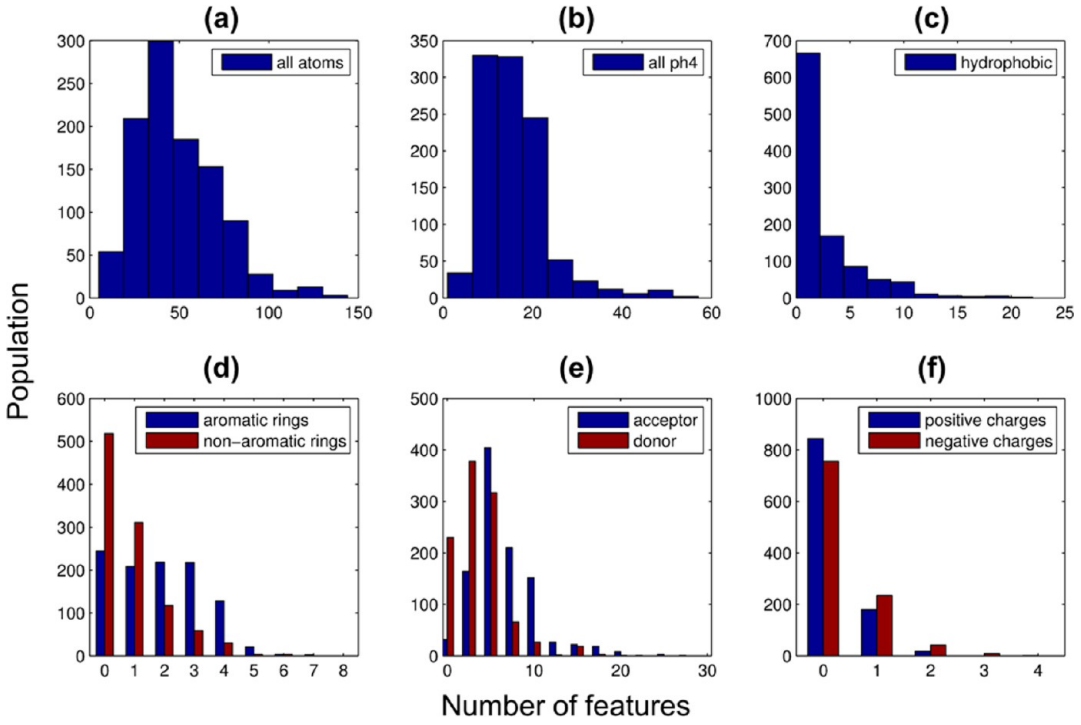


Figure 3. Number of pharmacophore features computed by DOCK FMS scoring function using the SB2012 docking testset (N = 1043 molecules).

from 1043 molecules in their X-ray pose taken from the SB2012 testset used in this work to gauge pose reproduction and crossdocking accuracy. As a reduced representation, the pharmacophore model derived for each molecule contains (on average) a much smaller number of pharmacophore points (16.0) relative to the total number of atoms (49.2) as shown in Figure 3 (panel b vs a). In terms of specific features, molecules in SB2012 contain on average of 1.9 aromatic rings, 0.9 nonaromatic rings, 2.6 hydrophobic groups, 3.6 hydrogen bond donors, and 4.6 hydrogen bond acceptors. Values for the two latter features are indicative of the druglike characteristics of many of the compounds in SB2012 for which ∼80% have less than 5 hydrogen bond donors and ∼58% have less than 10 hydrogen bond acceptors in rough agreement with Lipinski-like[34] rules. About 1/4 of the testset contains molecules with positively (199) or negatively (287) charged functionality.

In principle, given the smaller feature space, use of pharmacophore models should yield faster run times than an all-atom based scoring function. In terms of rescoring poses without sampling, timing tests indicate that under the current conditions, computing the pharmacophore matching similarity (FMS) score between two molecules is faster than computing the standard energy score by about 3.5 fold. Comparing production times when using the FMS method to drive ligand sampling is less straightforward, due to the much larger numbers of poses generated when using FMS compared to SGE (discussed further below). However, when normalized by the size of the final pose ensemble retained using FMS or SGE methods, time per pose with FMS is faster by about 1.5 fold.

2.2. Pharmacophore Matching Similarity (FMS) Scoring Function. After computing the pharmacophore (ph4) model using the protocol described above for both the reference and candidate poses, molecular similarity between the two poses is evaluated by the degree of pharmacophore overlap, termed here pharmacophore matching similarity score (FMS score). For each pharmacophore point A with pharmacophore label a, Cartesian coordinate $\vec{x} = [x_1, x_2, x_3]$ and direction vector $\vec{v} = [v_1, v_2, v_3]$ in the reference pharmacophore, is compared to every pharmacophore point B_i in the candidate pharmacophore in three steps: (i) label check, (ii) distance check, and (iii) direction check. The pharmacophore label a is used to eliminate pharmacophore points in the candidate pharmacophore that have different labels. The distance between A and the candidate pharmacophore point B_i, computed as $d_i = \|\vec{x} - \vec{y}_i\| = [\Sigma_{j=1}^{3}(x_j - y_j^i)^2]^{1/2}$ where $\vec{y}_i = [y_1^i, y_2^i, y_3^i]$ is the Cartesian coordinate of B_i, is compared to a distance cutoff r. Only when $d_i \leq r$ will the corresponding pharmacophore point B_i be further investigated. A constant r value is assigned to all reference pharmacophore points as a default parameter, but for a ring (aromatic or nonaromatic) the radius of the ring is assigned as r. The scalar projection of the normalized direction vector $\vec{v}$ onto that of B_i, $\vec{w}_i = [w_1^i, w_2^i, w_3^i]$ is calculated. The condition that the vector projection $\vec{v} \cdot \vec{w}_i = \Sigma_{j=1}^{3} v_j \times w_j^i \geq \sigma$ implies the angle between the two direction vectors $\vec{v}$ and $\vec{w}_i$ is within arccos σ, which ensures that the two vectors are pointing in approximately the same direction. A perfect vector overlap (when $\vec{v} = \vec{w}_i$) between two normalized direction vectors will be $\vec{v} \cdot \vec{w}_i = \|\vec{v}\| = 1$. By default, a scalar projection cutoff of $\sigma = \cos(45°) \approx 0.7071$ is used. Note that for hydrophobic and charged feature labeled points, $\vec{v} \cdot \vec{w}_i \geq \sigma$ is always true, as the same default value of (1,0,0) is assigned to both $\vec{v}$ and $\vec{w}_i$. For a ring, the absolute value of the scalar projection $|\vec{v} \cdot \vec{w}_i|$ is used to account for its orientation (i.e., vectors above and below the plane of the ring). If all of the above criteria are met then the two pharmacophore points A and B_i are deemed a match.

In Figure 4, three ARO pharmacophore point pairs are shown to illustrate how the three criteria (label, distance, and direction) are used to identify matches in rings. The first criterion (same label) is met by all three pairs as the

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

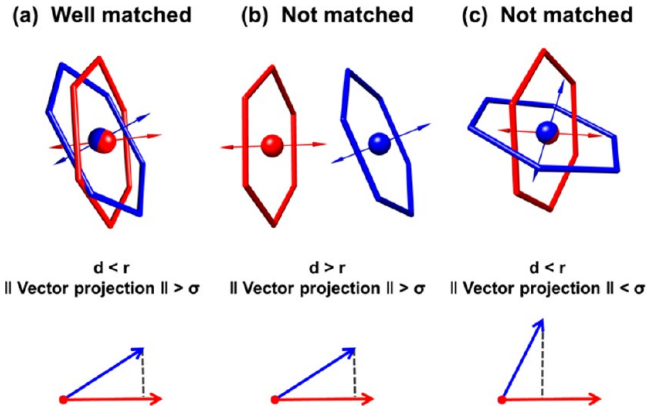


Figure 4. Example pharmacophore matches for aromatic rings showing: (a) well-matched case with the same labels, small distance, and similar vector directions, (b) not matched case with the same labels, large distance, and similar vector directions, and (c) not matched case with the same labels, small distance, and different vector directions.

pharmacophore points shown are all labeled as aromatic rings (ARO). The first pair (Figure 4a) has both a small distance ($d \leq r$) and good directional agreement ($|\vec{v} \cdot \vec{w}_i| > \sigma$) and thus represents a well-matched case. The second pair (Figure 4b), although the ring vectors are well-aligned, is not matched due to the large distance between the pharmacophore centers. For the third pair (Figure 4c), although the distance between ring centers is small, this case is also not considered a match due to the large difference in ring vector orientation ($|\vec{v} \cdot \vec{w}_i| < \sigma$).

All matched point pairs between the reference and candidate pharmacophore models are investigated by their geometric relationships to obtain a quantitative measurement of matching. The residual between two matched points is defined as $\delta_A^i = [(d_i)^2/|\vec{v} \cdot \vec{w}_i|]^{1/2}$ which takes into account both the distance and overlap in direction. After comparing pharmacophore point A with all candidate pharmacophore points B_i, the best matched point B_+ with the lowest matching residual δ_A^+ will be retained for the pharmacophore matching similarity (FMS) score calculation. If no match was found for A then it will not contribute to the residual term of FMS score. The residual term in combination with a match rate term defines the numerical value of the FMS score via eq 1.

$$FMS = \begin{cases} k\left(1 - \dfrac{n}{N}\right) + \sqrt{\dfrac{\sum_{j=1}^{n} (\delta_{A_j}^+)^2}{n}} & , \ n > 0 \\ X & , \ n = 0 \end{cases} \quad (1)$$

Here, k is a constant parameter; n stands for the total number of matches (note that for each reference pose pharmacophore point, one match is counted at most); N is the number of pharmacophore points in the reference pharmacophore; and $\delta_{A_j}^+$ represents the best matching residual of a matched reference pharmacophore point A_j. On the basis of similarity measurements in graph theory,[35,36] the FMS score uses the match rate term $k(1- n/N)$ to prioritize poses with higher numbers of pharmacophore matches to the reference pose. Poses with similar numbers of matches will be differentiated by their root-mean-square matching residuals $[\sum_{j=1}^{i} (\delta_{A_j}^+)^2/n]^{1/2}$. Note that the total number of matches n needs to be larger than zero for eq 1 to give a reasonable value. When no match is identified ($n = 0$), an arbitrary large score X is assigned (X is set to be larger than the upper bound of FMS score value when $n > 0$). For any reference and candidate pair of molecules, FMS score ranges between 0 (perfect match) and X, which depend on choices for k, distance cutoff r, and scalar projection cutoff σ. For pharmacophore-based docking, lower FMS scores are more desirable. Figure 5 outlines schematically the overall process using DOCK.

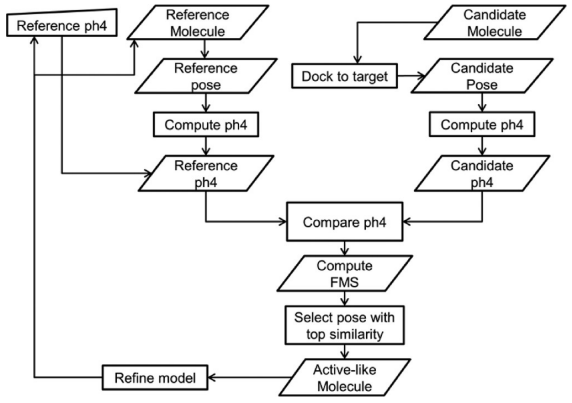


Figure 5. Flowchart schematic outlining pharmacophore-based virtual screening in DOCK.

To determine a default set of values for k, r, and σ in eq 1, we performed a series of rescoring tests using ligand geometries generated with the standard DOCK protocol, for comparison with crystallographic references, and pose reproduction success (defined in the next section) was determined. Four values for k (1, 2, 5, and 10), three values for r (0.5 Å, 1.0 Å, and 1.5 Å), and three values for σ (30°, 45°, and 60°) were examined. As a general rule, use of stricter matching criteria (shorter distance cutoff r, smaller angle cutoff σ) led to lower docking success rates. In addition, the success rate increased as the matching rate term weight k was increased from 1 to 5 but remained relatively steady from k = 5 to 10. With these results taken into consideration, the set comprising k = 5, r = 1 Å, σ = 45°, and X = 20 yielded generally good pose reproduction success and had values which were roughly in-between the different ranges explored. Although other combinations might also have been suitable, this set was ultimately employed for all subsequent FMS sampling and scoring experiments used in this work.

3. VALIDATION METRICS AND COMPUTATIONAL DETAILS

3.1. Pose Reproduction Details. In order to approximate the accuracy of ligand poses predicted by a given protocol for unknown systems, pose reproduction control experiments are performed over a large number of crystallographic complex structures. Ideally, the best-scored docked pose should agree with the crystal pose. Following our previous work,[27] docking results are categorized as one of three outcomes: docking success (success), scoring failures (score fail), and sampling failures (sample fail). Over a large data set, the percentage of success + score fail + sample fail =100%. Docking success is defined when the RMSD between the best scored pose and native (crystal) pose is ≤2 Å. A scoring failure is defined when a close-to-native pose is sampled, but the best scored pose is >2 Å from the native pose. Finally, a sampling failure is defined if none of the sampled poses are within 2 Å of the native pose.

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

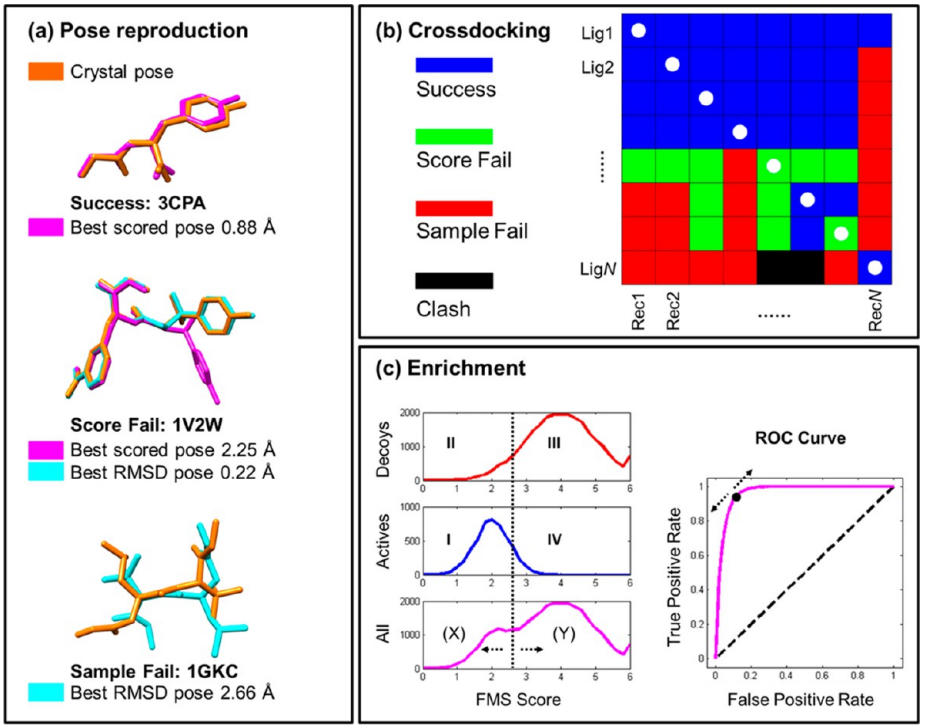


Figure 6. Validation metrics used to evaluate DOCK scoring functions. (a) Pose reproduction cases with different outcomes: Success (top, PDB code 3CPA), Score Fail (middle, PDB code 1V2W), and Sample Fail (bottom, PDB code 1GKC). Crystal poses in orange, best scored poses in magenta, best RMSD pose in cyan. (b) Representative crossdocking heatmap showing docking outcome as a function of docking all ligands (Lig1, Lig2, ..., LigN) to all receptors (Rec1, Rec2, ..., RecN) for an aligned group of proteins with nearly identical sequence homology. (c) Hypothetical database enrichment results showing a partitioning of data based on FMS score ranking (0 to 6) for a group of ligands (left bottom, magenta curve) comprised of a known active ligand set (left middle, blue curve) and inactive decoy set (left top, red curve). The vertical dashed line represents a hypothetical FMS score cutoff dividing the total group into (*X*) predicted positive and (*Y*) predicted negative sets which can be partitioned into four quadrants (I–IV) defined respectively as true positives (TP, I), false positives (FP, II), true negatives (TN, III), and false negatives (FN, IV). Also shown is an ROC curve, which for this example plots individual points which correspond to various FMS score cutoffs in the left panel. The coordinate of each point is determined by the false positive rate and true positive rate at that FMS score cutoff.

Representative visual examples of the three outcomes are shown in Figure 6a. For ligands of druglike size, low RMSD values also typically correspond to good visual overlap between docked and reference ligand poses. All statistics reported in this work make use of "symmetry corrected" RMSDs to account for chemically identical functionality (i.e., symmetric ring flips, carboxylate flips, etc.) or completely symmetric molecules, adopting visually indistinguishable conformations as described in detail previously.[37] The updated pose reproduction database termed SB2012 (an update of the SB2010 database)[27] was used for all pose reproduction and crossdocking (defined below) experiments. The set, derived from complexes in the protein databank (PDB), contains 1043 protein–ligand systems in ready to DOCK format and is freely available online at www.rizzolab.org.

All DOCK experiments in this work employed well-defined receptor and ligand setup protocols, in conjunction with the flexible ligand sampling protocol termed FLX, as previously described.[27] Briefly, in terms of receptor setup, several accessory programs are used to compute a molecular surface (DMS),[38] generate docking spheres to guide sampling (SPHGEN),[39] and precompute the potential energy on a grid which speeds up the docking calculations (GRID).[40] Key setup parameters include the use of 6–9 Lennard-Jones and distance dependent dielectric (ε-4r), a 0.3 Å resolution, and a grid box size extending 8 Å in all directions based on the docking spheres (75 spheres max). Key docking parameters include use of the on-the-fly anchor-and-grow algorithm to orient and assemble ligands layer-by-layer, retaining a maximum of 5000 completely grown conformers to be ranked by the primary scoring function and saving a maximum of 100 conformers (after clustering to remove redundancy, RMSD $\leq$ 2 Å). Ligands were energy-minimized at each stage of conformational search (500 iterations per cycle per anchor/step max), and those exceeding a total score cutoff of 100.0 were removed.

The different functions employed in this work include: (1) single grid energy (SGE) score, (2) DOCK Cartesian energy (DCE) score which is equivalent to SGE but in Cartesian space, (3) pharmacophore matching similarity (FMS) score, and (4) the combination of the two-termed FMS+SGE (or FMS+DCE) score. For the combined function, the FMS score was weighted by 10-fold so when summed together the FMS and SGE (or DCE) terms would be more equally balanced.

3.2. Crossdocking Details. In addition to pose production experiments, crossdocking was employed in which highly homologous protein complexes, with nearly identical structure and sequence (termed here a protein receptor family), are aligned into a common reference frame and each ligand is docked into each receptor as shown in Figure 6b. Such families inherently contain variability due to different crystallization conditions, cocrystallization with different ligands, as well as receptor point mutations, among others. Nevertheless, the hypothesis in crossdocking is that ligands should adopt similar binding geometries in highly homologous receptors, provided

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083–1102

there are no large deformations in the binding site or incompatible mutations. The results are expressed as a $N \times N$ heatmap (N = number of systems) with docking success plotted in blue, sampling failures plotted in red, and scoring failures plotted in green (Figure 6b). As before, a 2 Å RMSD cutoff is used to evaluate success. The diagonal elements (Figure 6b, white dots) represent cognate protein−ligand pairs and thus represent experimental references. Off-diagonal elements are "theoretical" protein−ligand pairs and the reference, in some instances, may be incompatible. To identify incompatible elements, we employ a clash matrix check,[27] independent of the actual crossdocking experiment, in which all matrix complexes (representing cognate and theoretical references) are subject to a short restrained energy minimization. If the minimized ligand pose moves >2 Å from the starting pose, or the pose bears an unfavorable energy score (>0 kcal/mol), the specific reference pair containing the clash is not included is crossdocking success evaluations (Figure 6b, black squares). All crossdocking studies employed the FLX docking protocol, and results are reported for both the diagonal and the entire matrix.

3.3. Enrichment Details. A third method used to evaluate docking methods is enrichment (Figure 6c). Databases such as the directory of useful decoys (DUD)[41] and the newer enhanced version called DUD-E[28] contain large sets of known active compounds (and property-matched decoys), which are docked to a specific target and the results are rank-ordered. Good enrichment is achieved when greater numbers of actives are ranked earlier in the list compared to the decoys. For more in-depth discussion on using DOCK to estimate enrichment, interested readers should consult Brozell et al.[42] Briefly, for this work, ranked results were visualized as receiver operating characteristic (ROC) curves which plots how the true positive rate (true positive/total positive) changes relative to the false positive rate (false positive/total negative). Accompanying area under the curve (AUC) analysis was also performed and used to estimate fold enrichment values (FE = AUC/AUC_{random}), relative to random, at 0.1%, 1%, 10%, and 100% of the database examined. For virtual screening, early enrichment is of particular importance, as typical applications will only focus on (i.e., purchase) small subsets of molecules ranked very early (i.e., 0.1−1%) in the database. In the theoretical example shown in Figure 6c, which employed FMS score to rank active and decoy ligands shown in the left panels, the ROC curve on the right represents a good enrichment case relative to random (Figure 6c, magenta vs dashed line). By specifying a specific score cutoff (Figure 6c left bottom panel, dashed line) the data can also be partitioned into two groups for which molecules with smaller scores (better overlap) are defined as predicted positives (X) and molecules with higher scores (worse overlap) defined as predicted negatives (Y). If, as in the present example, the results are in fact known, this allows ligands in the active group to be classified as true positive (I) or false negative (IV), and ligands in the inactive (decoy) group classified as false positive (II) and true negative (III). By varying the cutoff, the number of molecules in the four subsets I−IV will change accordingly.

Enrichment studies employed the 15 DUD-E systems shown in Table 3.[28] The receptor PDB files were already available in SB2012 (same PDB code as DUD-E), and the active and decoy ligands were downloaded from the DUD-E Web site and used as is. It is important to note that some ligands (active and decoys) for these systems contain multiple entries representing, for example, different tautomers or protonation states. For all enrichment analyses, in the case of duplicate id codes, only the best-scored molecule was retained. For each system, the native cognate ligand in the original PDB file is used as the pharmacophore reference for FMS scoring. As in the pose reproduction and crossdocking studies, the enrichment tests also employed the FLX docking protocol. With this protocol, predicted ligand poses with accompanying scores were obtained for approximately all but 2% of the actives and decoys listed in Table 3.

Table 3. Systems Used for Enrichment Tests

PDB	system	#actives[a]	#decoys[a]	description
2HZI	abl1	295	10885	tyrosine-protein kinase ABL
1E66	aces	664	26373	acetylcholinesterase
2VT4	adrb1	458	15958	beta-1 adrenergic receptor
1L2S	ampc	62	2902	beta-lactamase
1BCD	cah2	835	31710	carbonic anhydrase II
1R9O	cp2c9	183	7574	cytochrome P450 2C9
2RGP	egfr	832	35442	epidermal growth factor receptor erbB1
1SJ0	esr1	627	20818	estrogen receptor alpha
3CCW	hmdh	299	8884	HMG-CoA reductase
1UYG	hs90a	125	4942	heat shock protein HSP 90-alpha
2AA2	mcr	193	5240	mineralocorticoid receptor
1KVO	pa2ga	127	5216	phospholipase A2 group IIA
2GTK	pparg	723	25867	peroxisome proliferator-activated receptor gamma
1NJS	pur2	201	2725	GAR transformylase
1C8K	pygm	114	4045	muscle glycogen phosphorylase

[a]Systems taken from DUD-E database.[28]

4. RESULTS AND DISCUSSION

4.1. Pose Reproduction Results. Table 4 shows pose reproduction outcomes computed for the three DOCK protocols tested (SGE, FMS, and FMS+SGE), in which a given function was used for both sampling and scoring (diagonal blocks in gray box) or when rescored using the other two scoring functions (off-diagonal blocks). All experiments were performed under the same conditions except for the sampling and/or scoring method employed. It is important to emphasize that use of an alternative function to rerank an ensemble of poses generated by any given method (Table 4, off-diagonal blocks) will, in most cases, lead to a different group of top-scored results, but the number of sampling failures remains unchanged.

In general, the diagonal results (Table 4, gray boxes) using the three different methods yield high percentages of success across the 1043 systems in SB2012 with the FLX ligand protocol. Importantly, the SGE success rate (72.5%) is consistent with earlier work from our laboratory,[7] using a smaller data set (68.5%, N = 780), indicating good reproducibility of DOCK. Overall, the diagonal results in Table 4 reveal a clear trend in terms of outcome with success following SGE < FMS < FMS+SGE and sampling and scoring failures following SGE > FMS > FMS+SGE. The very high success rates when using FMS (93.5%) or the combination FMS+SGE (98.3%) is significant and represents a 20−25% improvement over the standard DOCK method employing SGE (72.5%). On one hand, such high success rates are expected given that for any system the X-ray reference ligand and docked ligand are the same molecule in terms of topology

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

Table 4. Pose Reproduction Results Employing SGE, FMS, and FMS+SGE Scoring Functions

		Sampling[a]					
Scoring	Outcome	SGE		FMS		FMS+SGE	
SGE	Success	756	72.5%	610	58.5%	854	81.9%
	Score Fail	185	17.7%	394	37.8%	182	17.4%
	Sample Fail	102	9.8%	39	3.7%	7	0.7%
FMS	Success	860	82.5%	975	93.5%	1035	99.2%
	Score Fail	81	7.8%	29	2.8%	1	0.1%
	Sample Fail	102	9.8%	39	3.7%	7	0.7%
FMS+SGE	Success	876	84.0%	719	68.9%	1025	98.3%
	Score Fail	65	6.2%	285	27.3%	11	1.1%
	Sample Fail	102	9.8%	39	3.7%	7	0.7%

[a]SGE sampling size = 89083 poses, FMS sampling size = 337674 poses, FMS+SGE sampling size = 59237 poses.

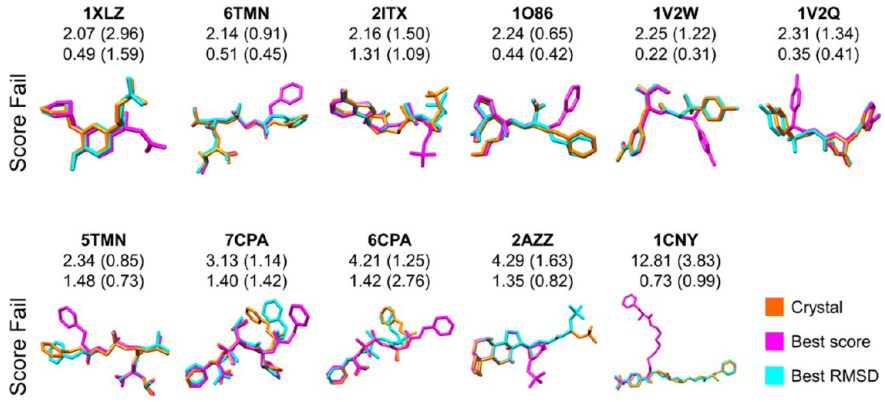


Figure 7. Eleven scoring failures derived from FMS+SGE guided docking showing overlaid poses, PDB code identifier. RMSD in angstroms and FMS scores in parentheses for the best FMS+SGE scored pose (first row, magenta) and the best FMS+SGE RMSD pose (second row, cyan) relative to the crystal pose in orange.

and thus have the exact same number of pharmacophore features. In actual practice, for virtual screening, the number of features between a reference and candidate would change as each new ligand was docked. Nevertheless, the good correspondence in these validation tests provides strong evidence the newly implemented DOCK pharmacophore labeling, modeling, and overlap routines are behaving as expected and yield robust results over a large pose reproduction testset. Importantly, the FMS method is straightforward to use and only requires that the user input a reference molecule consisting of a single 3D conformation. The processing of the candidate pose(s) to determine FMS scores is done automatically and on-the-fly. Ongoing work to allow a text-based pharmacophore reference to be used as a query will further simplify the procedure of customizing inputs for FMS score calculation.

Systems with Failures. Of the three methods tested, the FMS+SGE protocol yields the lowest sampling (0.7%) and scoring (1.1%) failure rates on the diagonal. In an attempt to understand what led to the small subset of failures (N = 18), docked poses for the group were examined. Out of the seven sampling failures, one system did not complete growth, which, although infrequent, can happen using DOCK under some circumstances. And for the remaining 6 sampling failures, 4 are relatively large molecules with up to 35 rotatable bonds and thus extremely challenging for any docking protocol. In terms of the 11 scoring failures, a noteworthy result (Figure 7) is that 7 out of the 11 systems (PDB codes: 1XLZ, 6TMN, 2ITX, 1O86, 1V2W, 1V2Q, and 5TMN) actually show good correspondence both in terms of visual overlap as well as RMSD (2.07−2.34 Å). Thus, these 7 can be classified as "near misses" for which only a part of the ligand geometry adopts a conformation different than the X-ray pose. Consistent with expectation, in all but two cases (7CPA, 6CPA), geometries corresponding to the best RMSD also have a lower FMS score. The fact that the FMS+SGE protocol correctly identifies a nativelike pose in nearly all 1043 cases is notable.

Rescoring. In terms of the off-diagonal blocks (Table 4), rescoring the standard SGE results (72.5%) with FMS (82.5%) or FMS+SGE (84.0%) reveals a similar trend with SGE < FMS < FMS+SGE as in the diagonal experiments. Here, as rescoring cannot "rescue" incorrectly sampled geometries, the maximum success rate attainable is a function of the poses originally sampled, which for SGE is 90.2% (e.g., 100% - 9.8% sampling failures). This specific experiment is important as the improvement in success when rescoring SGE-derived results with FMS or FMS+SGE (10−11%) suggests the current implementation is a viable way to postprocess docked poses and identify those compounds with good pharmacophore overlap to a reference. This procedure would be a particularly useful tool to aid virtual screening as discussed further below. Rescoring results for the group derived from FMS+SGE sampling shows similar results, with FMS (99.2%) yielding a significantly higher success rate than SGE (81.9%).

The most dramatic changes in terms of pose reproduction involve using SGE (58.5%) or FMS+SGE (68.9%) to rescore

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

the pose ensembles derived from FMS-only sampling (93.5%). These reduced success rates likely stem from the fact that the FMS score accounts only for overlap between pharmacophore features derived from the reference ligand structure and the receptor is "invisible" during sampling. The end result is that poses generated using FMS alone may clash with the target protein when rescored in "energy space" despite high pharmacophore overlap. However, as the pairing of energy and pharmacophore overlap (FMS+SGE) leads to relatively high success rates when rescored in SGE-space, as noted above, the combined function is likely to be preferred when a receptor structure is available. Nonetheless, the 58% success rate obtained with SGE rescoring can be considered encouraging considering that ligand sampling with the anchor-and-grow algorithm was done in the absence of a receptor. Thus, for ligand-only based design, the FMS protocol appears to be capable of enriching for energetically favorable poses by matching only to a reference pharmacophore. The caveat of course is identifying suitable pharmacophores in the absence of crystallographic information.

Ensemble Properties. A protocol designed to enrich for ligands with poses close to a native structure should, in theory, yield favorable scores using any reasonable scoring function. To examine in more detail how properties of molecules generated with one protocol may differ when rescored with another, histograms of the resultant SGE and FMS scores were plotted using each of the three different pose ensembles obtained with SGE, FMS, or FMS+SGE methods. As expected, and consistent with the rescoring results in Table 4, use of the FMS function alone to derive poses does lead to overall less favorable DOCK energies (Figure 8 top, red) when rescored in SGE-space

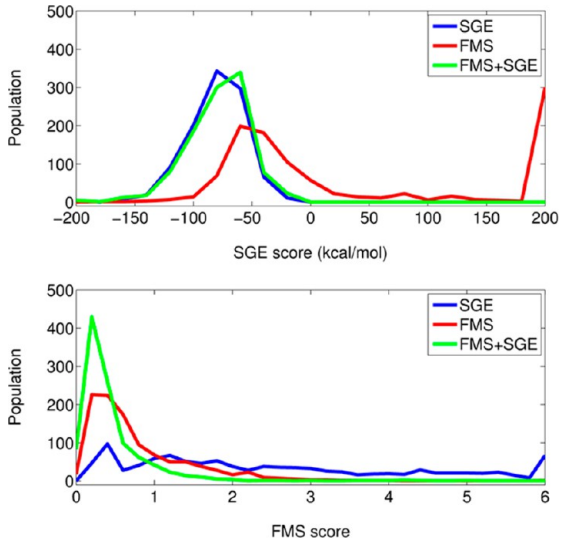


Figure 8. SGE (top) and FMS (bottom) score histograms using ensembles derived from SGE (blue), FMS (red), or FMS+SGE (green) driven sampling methods.

compared to FMS+SGE (Figure 8 top, green) or SGE (Figure 8 top, blue). The large positive peak at 200 kcal/mol (Figure 8 top, red) represents those systems for which large positive energies were obtained due to geometric clashes occurring between ligand and protein. However, an encouraging number of the poses derived from FMS sampling do yield favorable energies. At first glance, the fact that the SGE and FMS+SGE energy histograms (Figure 8 top, blue and green) are nearly superimposable is somewhat surprising, especially considering the two ensembles yield substantially different success rates (SGE = 72.5% vs FMS+SGE = 98.3%). However, given the underlying complexity of binding energy landscapes, ligand poses with distinctly different binding geometries may in fact yield similar energy scores (and vice versa), thus the observed SGE overlap in Figure 8 (top panel) is not unreasonable.

As shown in Figure 8 (bottom), FMS score distributions show much greater separation, indicating greater sensitivity in contrast to the SGE score distributions shown in Figure 8 (top). Here, SGE sampled poses yield a much wider almost uniformly distributed range of FMS scores (Figure 8 bottom, blue) compared to FMS (Figure 8 bottom, red) or FMS+SGE (Figure 8 bottom, green) sampled poses which have large peaks around 0.5, indicative of high pharmacophore overlap. Importantly, the FMS+SGE combination containing both geometric and energetic components to guide growth yields energy scores on par with standard SGE-guided docking poses (Figure 8 top, green vs blue) and matches the pharmacophore models even better than FMS-only docking (Figure 8 bottom, green vs red).

Ensemble Sizes. An additional interesting observation from the results in Table 4 is the larger number of final docked poses obtained using FMS (337674) compared to SGE (89083) or FMS+SGE (59237). The much larger ensemble generated with FMS corresponds to an increase in total docking time, which could be of concern, although when normalized by the number of poses kept, the FMS function is actually faster than SGE by about 1.5 fold. The most likely explanation for the increased size involves reduced pruning. Current experiments employed a standard DOCK input file specifying a maximum score cutoff of 100.0, larger than the upper bound of the FMS function [0, 20]. Thus, poses are not as vigorously pruned during growth compared to protocols that employ energy-based functions (themselves not bounded). The significantly larger ensemble from FMS-sampling also likely contributes to the reduction in docking success rate associated with SGE rescoring because of the greater number of alternative (decoy) poses associated with system. Future studies to optimize the maximum score cutoff parameter would be worthwhile.

4.2. Quadrant Partitioning Using FMS Score. Although no score cutoffs were used to define success in the pose reproduction tests in Table 4, if both a RMSD cutoff and score cutoff are defined then the results can be classified in one of four different quadrants (see Figure 9b) defined as (I) true positive (TP), good score and low RMSD; (II) false positive (FP), good score and high RMSD; (III) true negative (TN), bad score and high RMSD; (IV) false negative (FN), bad score and low RMSD. To highlight properties of the new DOCK pharmacophore function, Figure 9 focuses on the results derived using only the FMS-guided sampling protocol discussed above (success = 93.5%, sampling failure = 3.7%, scoring failure = 2.8%). Dashed green lines at RMSD = 2 Å and FMS = 2 delineate the four quadrants.

Figure 9a plots the large "all poses" set consisting of 239486 ligand conformations with FMS < 20 out of the total sampled space obtained with FMS sampling (337674 poses). Here, the small separate cluster of points located in the TP region (lower left quadrant), which shows roughly linear correlation with RMSD, corresponds to mostly docking successes compared to the highly populated TN region (upper right quadrant) containing many thousands of points for which the correlation between RMSD and FMS begins to diverge as FMS values

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

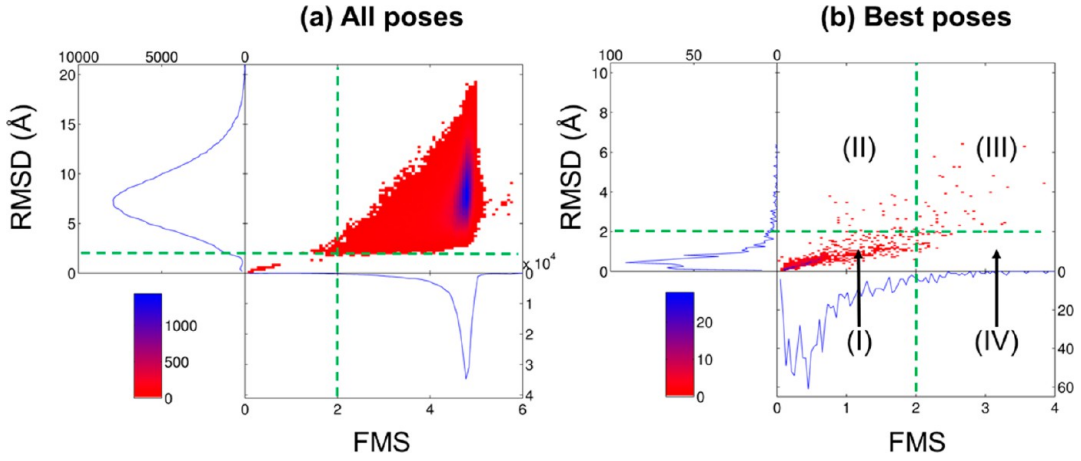


Figure 9. 2D Histograms of FMS score and RMSD for (a) all poses (N = 239486) and (b) best scored poses (N = 1041) generated using FMS guided sampling of 1043 systems. Poses without matches (FMS = 20) not included in histograms. Color reflects density (population).

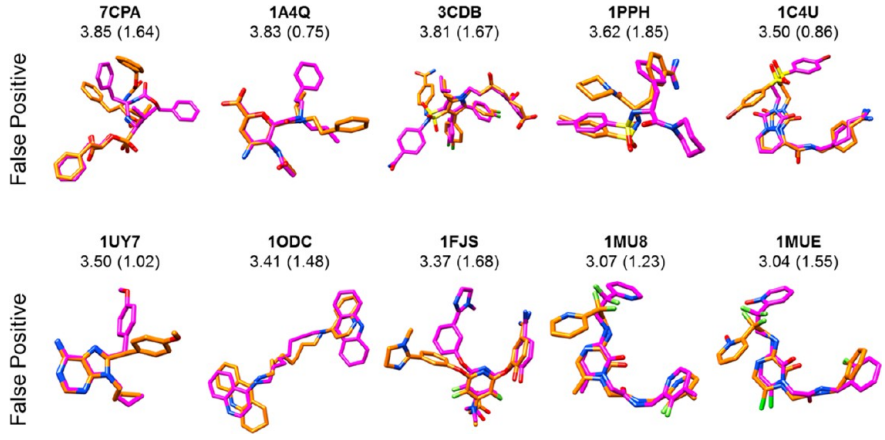


Figure 10. Ten out of twenty-eight FP poses derived from FMS-guided docking with the largest RMSD values. Crystal poses in orange, best scored poses in magenta. RMSD in angstroms and FMS scores in parentheses.

increase. Unlike the standard SGE function, which typically shows little correlation with RMSD, the FMS method behaves more like RMSD given the geometric nature of the function. Importantly, the results in Figure 9a indicate that the FMS protocol is not only able to identify close-to-native ligand conformation with favorable scores (region I) but also correctly characterizes poses that are geometrically different from the reference by assigning unfavorable scores (region III).

Figure 9b plots the "best poses" set consisting of 1041 ligand conformations (1 system failed to dock, 1 system with FMS = 20 for the best score pose). As in Figure 9a, poses in the TP region again show roughly linear correlation with RMSD. In this case, however, as only a single pose for each system is retained, unlike the "all poses" case, the TN region is sparse. Ideally, a good function should maximize TP and minimize FP. With the present RMSD (2.0 Å) and FMS (2) cutoffs, 949 points are classified as TP and 26 are classified as FN. The remaining points (1042 − 975 = 67) are divided into 39 TN cases and 28 FP cases. Overall, the 97.3% TP rate (949/975) and 41.8% FP rate (28/67) indicates good quadrant partitioning. And, as expected, use of a smaller score cutoff will yield a reduction in TP but an improvement in FP. For example, use of an FMS cutoff = 1.5 yields a TP rate = 91.4% and a FP rate = 20.9%, and use of an FMS cutoff = 1.0 yields a TP rate = 78.1% and FP rate = 9.8%. As a point of comparison, comparable analysis by Balius et al.[7] for a similar TP rate = 79.8% yielded a higher FP rate = 46.2% using DOCK's footprint similarity method with a 0.6 score cutoff (based on normalized Euclidean distance) and 2 Å RMSD cutoff across 780 protein−ligand systems. In practice, the optimal choice of a numerical value for score cutoff to employ in a study to yield compounds with the desired properties is system-dependent. For example, in typical virtual screening applications, FMS score between candidate compounds and a reference would be expected to be higher (i.e., less overlap) than under the present pose reduction tests which compare compounds with identical topologies but different conformations.

False Positive (FP) Cases with FMS. While FMS in general yields excellent quadrant partitioning, an examination of the results was undertaken to determine the underlying cause of FP and FN classifications. Focusing on results from the "best poses" set (Figure 9b), Figure 10 presents the ten out of twenty-eight FP results (RMSD > 2 Å, FMS ≤ 2) with the highest RMSD. Analogous to that observed with the FMS+SGE scoring failures (Figure 7), FP poses derived with FMS-sampling show, for the most part, remarkably high overlap except for one end of the molecule. And in all ten cases, the poorly overlapped groups contain rings, which are weighted heavier by the RMSD function than FMS. System 1ODC is a particularly interesting case. Here, the ligand pose is semi-

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

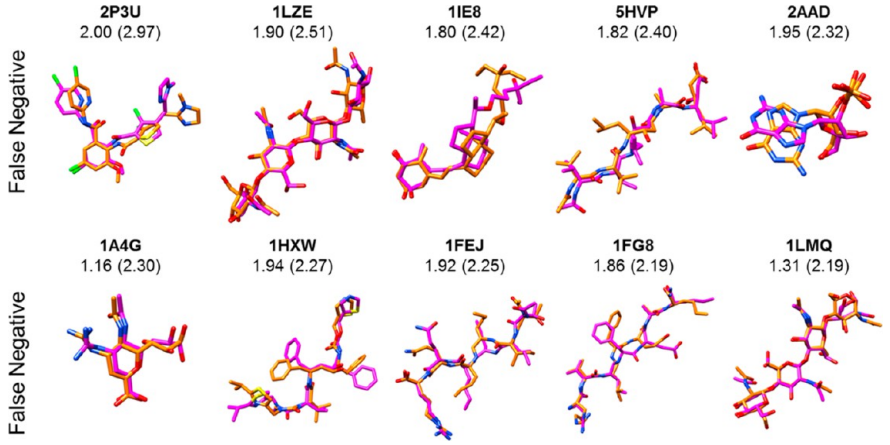


Figure 11. Ten out of twenty-six FN poses derived from FMS-guided docking with the largest FMS scores. Crystal poses in orange, best scored poses in magenta. RMSD in angstroms and FMS scores in parentheses.

symmetric and flipped by ca. 180° relative to the reference (magenta vs orange), resulting in overlap between two rings on one end with three rings from the other. Although the Hungarian algorithm used here in DOCK[37] to compute symmetry-corrected RMSD effectively accounts for the swap of functionality having identical chemical properties, the resultant value of 3.41 Å is still classified as a failure, largely as a result of one ring on either end (8 atoms total) not being matched. In contrast, the FMS score not only accounts for the symmetry but the good overlap between four out of six ring centers (and associated vector directions), which leads to a relatively low FMS score of 1.48. Overall, visual examination of these ten worst FP cases reveals a significant amount of physically reasonable matches and minimal mismatch and the classification of poses to this quadrant is, in most cases, understandable.

False Negative (FN) Cases with FMS. In terms of the FN examples (RMSD < 2 Å, FMS > 2), Figure 11 presents the ten out of twenty-six poses with the highest FMS scores. Immediately obvious compared to the FP examples is that the molecules here contain fewer aromatic rings, for the most part are larger and more extended, and have a higher number of more loosely matched hydrogen-bonding functional groups (most polar atoms in the FP cases are either tightly matched or not matched at all). This latter point is particularly important as relatively small changes in position of a hydrogen-bonding functional group can lead to relatively large changes in FMS overlap but minor effects on RMSD which is computed using only heavy (non-hydrogen) atoms. Although our standard preparation protocol for FMS scoring employs an energy minimization step to relax any hydrogen atoms added to the system, the positions adopted as a result of ligand sampling during growth may result in the candidate and reference poses having different hydrogen directions. This result highlights the need for care when preparing a molecule to be used as a "reference" for scoring candidate compounds. Despite being a distinctly different type of function, a similar conclusion was reached by employing the DOCK footprint function.[7] Despite this sensitivity, however, most of the FN cases have scores close to 2 that could easily be rescued by a minor increase in FMS cutoff to 2.5.

4.3. Crossdocking Results. In addition to pose reproduction, crossdocking experiments are a useful way to determine if different protocols can reproduce nativelike poses when ligands are docked to highly homologous protein binding sites from different crystallographic structures (see Figure 6b). Figure 12 displays outcomes across six protein families: carbonic anhydrase (CA, $N = 29$), carboxypeptidase A (CPA, $N = 8$), epidermal growth factor receptor (EGFR, $N = 15$), thermolysin (THERM, $N = 26$), HIV protease (HIVPR, $N = 60$), and HIV reverse transcriptase (HIVRT, $N = 21$). For comparison, both the diagonal (cognate protein−ligand pairs) and the entire matrix (all combinations) are shown. As before, three docking protocols were tested (SGE, FMS, and FMS +SGE). As shown in Figure 12a, this is a particularly challenging group of proteins with the standard SGE protocol yielding low diagonal successes (34.5−60.0%) for 5 out of 6 families. The exception is HIVRT for which the SGE success rate = 95.2%. In contrast, use of FMS (71.7−100.0%) or FMS +SGE (75.0−100.0%) yields significant improvement for cognate receptor−ligand pairs. Carbonic anhydrase is a particularly noteworthy example as the SGE diagonal success

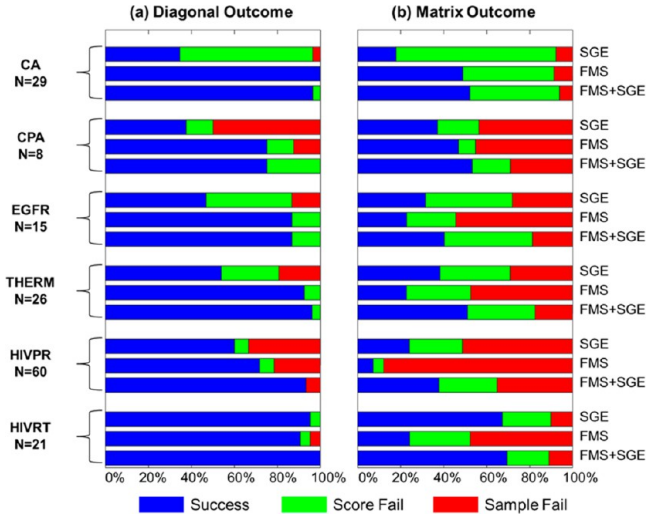


Figure 12. Crossdocking outcomes averaged across the diagonal (left) or total matrix (right) for six protein families: carbonic anhydrase (CA), carboxy peptidase A (CPA), epidermal growth factor receptor (EGFR), thermolysin (THERM), HIV protease (HIVPR), and HIV reverse transcriptase (HIVRT) using SGE (top), FMS (middle), and FMS+SGE (bottom) protocols. Success in blue, scoring failure in green, and sampling failure in red.

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

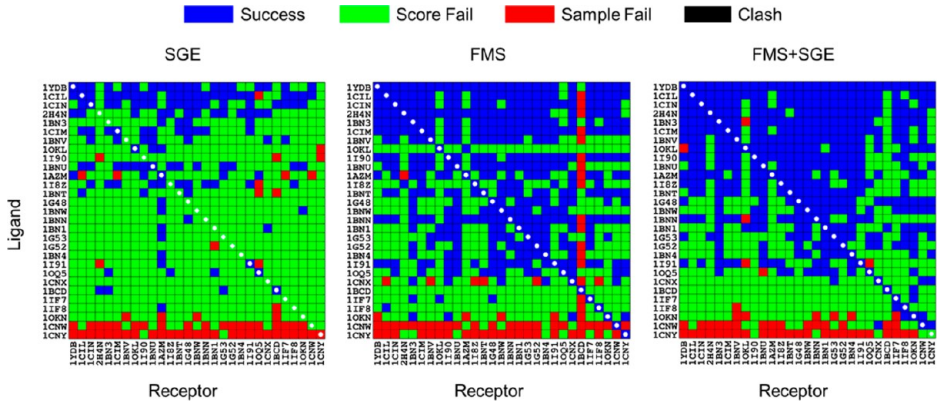


Figure 13. Crossdocking heatmaps using SGE, FMS, and FMS+SGE protocols for carbonic anhydrase (29 × 29 = 841 combinations).

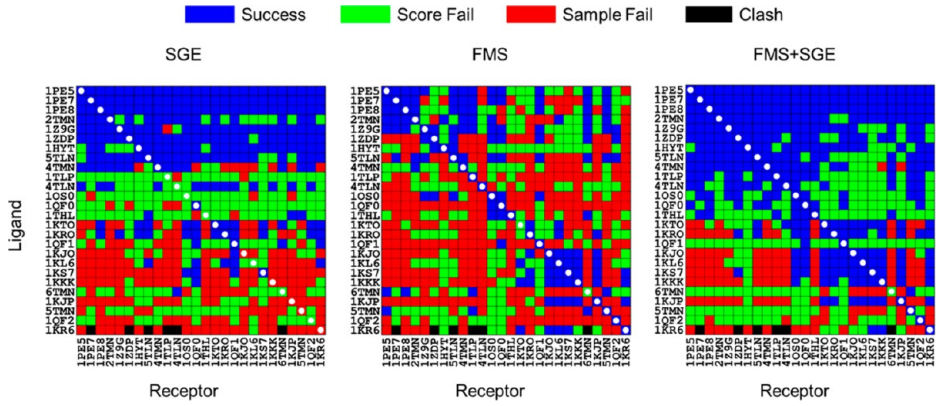


Figure 14. Crossdocking heatmaps using SGE, FMS, and FMS+SGE protocols for thermolysin (26 × 26 = 676 combinations).

increases from only 34.5% to near 100.0% using the FMS or FMS+SGE functions. Comparable enhancements in success for carbonic anhydrase were also reported by Balius et al.[7] when using the DOCK footprint similarity scoring function (82.8%) compared to SGE (31.0%).

As expected, for more challenging crossdocking experiments, matrix success (Figure 12b) using any of the scoring functions are in general significantly lower than their diagonal counterparts (Figure 12a). As a baseline, use of SGE yields an averaged matrix success of 36.0% compared to the diagonal at 54.6%. In contrast to the diagonal results, interestingly, use of FMS alone for crossdocking shows improvement over SGE in only two cases (CA and CPA). However, in all cases, the combined FMS +SGE function always yields a better matrix success than does SGE. Analogous to the diagonal results, the matrix outcomes (Figure 12b) similarly reveal that carbonic anhydrase has the lowest overall matrix SGE success rate (17.8%) which increased the most among all systems tested when using FMS (48.8%) or FMS+SGE (52.1%). Figure 13 compares the heatmaps for carbonic anhydrase, derived from three independent docking sets of size 29 × 29 = 841 combinations, using SGE, FMS, and FMS+SGE methods. The maps visually highlight that SGE failures are primarily due to scoring (green squares), pinpoint which specific systems are involved, and indicate that FMS and FMS+SGE protocols significantly improve docking outcomes (more blue squares).

Additionally, visible in the FMS heatmap for carbonic anhydrase (Figure 13, middle) is the appearance of previously unseen sampling failures specifically localized to column 1BCD. It is important to note that the RMSD calculations in both diagonal and off-diagonal experiments always involve compounds of the same topology. However, for pharmacophore overlap calculations involving off-diagonal elements, the pharmacophore reference and the candidate molecule being docked are usually of different topology. In such cases, FMS-guided docking may drive sampling in a direction that will not necessarily agree with the RMSD reference. Calculation of the pharmacophore overlap between all aligned crystallographic references for carbonic anhydrase indeed shows 1BCD has the poorest reference FMS scores (between the pharmacophore reference and the RMSD reference) when averaged across all columns (FMS = 5.15) or all rows (FMS = 5.51), which is appreciably above the overall average (FMS = 3.38) across all reference pairs. Inspection further revealed that the ligand from 1BCD has only one rotatable bond and a molecular weight of 148.1 g/mol, which is markedly smaller than the average ligand in this family with 5.1 rotatable bonds and a molecular weight of 339.7 g/mol. Thus, crossdocking of ligands to receptor 1BCD, using FMS alone, is not expected to be consistent with the 1BCD reference sampling space, which leads to the observed sampling failures.

Additionally, more dramatic examples of this phenomenon manifest themselves in the heatmaps for thermolysin as shown in Figure 14. Here, in contrast to carbonic anhydrase, crossdocking with SGE yields a higher overall success rate of 38.2% (Figure 14 left, blue) but with a higher percentage of sampling failures (29.1%, red). And, while the combined function FMS+SGE yields the overall best docking success rate (51.0%) for this family, use of FMS alone actually increases sampling failures (47.5%) relative to SGE (Figure 14 left vs

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

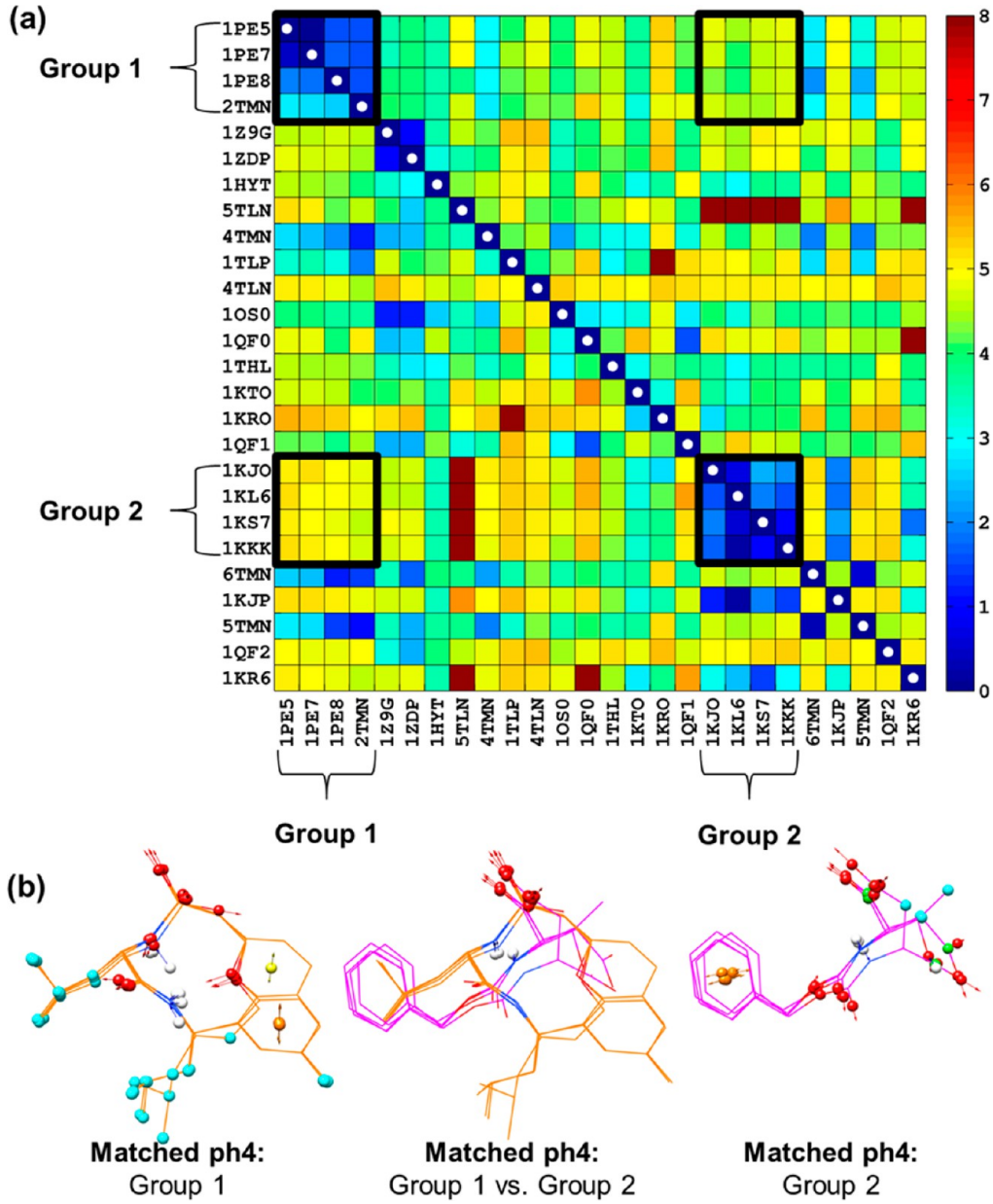


Figure 15. (a) FMS heatmap, using all crystallographic reference poses for thermolysin, with perfect overlap in dark blue (FMS = 0) and poorest overlap (FMS ≥ 8) in dark red. Group 1 submatrix defined by systems 1PE5, 1PE7, 1PE8, and 2TMN. Group 2 submatrix defined by systems 1KJO, 1KL6, 1KS7, and 1KKK. (b) Crystallographic reference overlays showing matched pharmacophore features for group 1 (left, orange), group 2 (right, magenta), and group 1 vs group 2 (middle).

middle, red) which, as described below, likely involves poor reference pharmacophore overlap. Close inspection of the crossdocking heatmaps reveals submatrices of size 4 × 4, defined here as group 1 (1PE5, 1PE7, 1PE8, 2TMN) and group 2 (1KJO, 1KL6, 1KS7, 1KKK), for which FMS sampling relative to SGE: (i) maintains docking success and/or (ii) rescues previously unsuccessful docking outcomes involving systems within the same group, and (iii) introduces docking failures for systems from different groups. To aid the discussion, Figure 15 shows a heatmap of FMS scores (as opposed to docking outcomes), derived from the X-ray references, with diagonal and off-diagonal submatrix blocks for groups 1 and 2 outlined as black boxes.

The FMS scores computed between all reference pairs indicate perfect overlap on the diagonal (FMS = 0, dark blue), but for the most part the majority of pairs have poor overlap (FMS = 3−8, green to dark red). A striking exception are the cases defined by groups 1 and 2 (Figure 15, black boxes) which all have relatively good reference FMS scores within the same group (two blue submatrices near the diagonal) but poor FMS scores between different groups (two green to yellow submatrices on the off-diagonal). This observation helps explain why FMS-guided docking yields 100% success across the submatrices formed within the same group (Figure 14, middle), but when using group 1 systems as a reference to guide docking of ligands in group 2, no matrix success is reported and only 1 success is obtained for the opposite case (other symmetric block). Structurally, the molecular cluster formed by ligands in group 1 occupies an extended space in the thermolysin binding pocket (Figure 15b, left) and contain

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

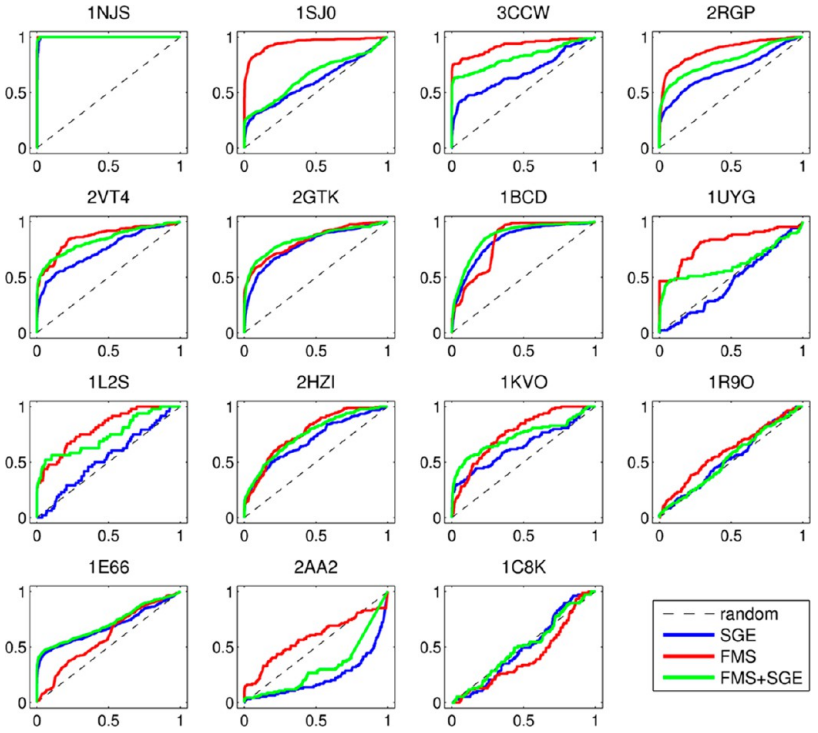


Figure 16. ROC enrichment curves for 15 DUD-E systems using SGE, FMS, and FMS+SGE protocol.

additional hydrophobic groups compared to group 2 (Figure 15b, right). Group 2 ligands cluster into a more slender volume anchored by an aromatic ring at one end and hydrogen bond acceptor on the other. As a consequence, groups 1 and 2 share only a few (1−3) matched pharmacophore points (Figure 15b, middle), which explains the poor FMS scores between off-diagonal reference ligands in addition to the poor docking outcomes. Interestingly, the addition of the energy term to the pharmacophore overlaps score (FMS+SGE score), using group 2 as a reference to dock group 1, yields 100% docking success. In contrast, using group 1 as reference to dock group 2 yields 100% sampling failure (Figure 14, right panel).

Finally, the overall poorest matrix success results using FMS (7.2%) or FMS+SGE (37.7%) docking is seen with HIVPR. Although high ligand flexibility is expected to play a role in the large number of sampling failures seen in the FMS matrix (Figure 12b, red) relative to other systems (31/60 of ligands have ≥15 rotatable bonds), the most likely cause is poor pharmacophore overlap between all pairwise combinations. Consistent with the discussions above, out of the 3600 pairwise combinations in the HIVPR crossdocking reference FMS matrix derived from crystallographic poses, only 220 pairs yielded reasonable pharmacophore overlap (FMS ≤ 3). In contrast, 2493 pairs have poor pharmacophore overlap (FMS ≥ 4.5) which, interestingly in this case, is about the same as the number of sampling failures (2816).

Overall, two key points have emerged from the current crossdocking studies: (1) FMS-guided success rates, particular for off-diagonal elements, are dependent on the similarity between the pharmacophore reference and the RMSD reference. (2) The FMS+SGE protocol generally improves crossdocking performance, relative to SGE or FMS, by integrating known binding profiles into the standard DOCK energy score.

4.4. Enrichment Results. Results for enrichment experiments, used to gauge how DOCK would perform in a virtual screening using SGE, FMS, or FMS+SGE protocols are shown in Figure 16 and Table 5. Receiver operating characteristic (ROC) curves and area under the curve (AUC) analyses were used to compute fold enrichment (FE = AUC_{curve}/AUC_{random}) values for docking active and decoy ligands taken from the DUD-E database.[28] For virtual screening applications, good early enrichment is considered to be critically important, thus FE was also computed at 0.1%, 1%, and 10% of the ranked database. For the current tests, the overall shape of the ROC curves vary from essentially perfect enrichment (1NJS) to random enrichment (1C8K) with most systems exhibiting good overall enrichment but with a visible dependence on which of the three docking functions was used. For the majority of systems, depending on which ROC region is examined, FMS (red curves) shows higher enrichment than SGE (blue curves), with FMS+SGE (green curves) being roughly in between (Figure 16). Across different ranges of the database, based on numerical AUC values, use of FMS or FMS+SGE consistently yield higher FE rates relative to SGE (Table 5). For example, at 0.1% of the database, 11/15 FE values using FMS and 11/15 FE values using FMS+SGE are enhanced relative to SGE (Table 5, column A). Similarly, at 1% of the database, 10/15 FE values using FMS and 13/15 FE values using FMS+SGE are enhanced relative to SGE (Table 5, column B). Comparable results are obtained at 10% and 100% of the database.

The fact that use of FMS+SGE yields generally lower enrichment outcomes than FMS is somewhat surprising given that FMS+SGE yielded higher success rates than FMS in pose reproduction experiments. However, it is important to note that the role of the SGE term in FMS+SGE is fundamentally different for pose reproduction given that different molecular conformers, as opposed to the different chemical species for enrichment, are what is rank-ordered. The most likely

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

Table 5. Fold Enrichment (FE) Results at Different Percentages of the Database (DB) Screened

		(A) FE @ 0.1% of DB[b]	(B) FE @ 1% of DB[b]	(C) FE @ 10% of DB[b]	(D) FE @ 100% of DB[b]
	random	1.00	1.00	1.00	1.00
system[a]	maximum	2000.00	200.00	20.00	2.00
1NJS	SGE	0.00	111.36	19.05	1.99
	FMS	1009.65	184.26	19.88	2.00
	FMS+SGE	0.00	150.85	19.52	2.00
1SJ0	SGE	88.96	22.09	4.98	1.21
	FMS	804.91	114.91	15.39	1.90
	FMS+SGE	382.87	48.69	5.90	1.30
3CCW	SGE	80.74	31.49	7.20	1.43
	FMS	1167.99	144.49	15.46	1.77
	FMS+SGE	932.51	116.07	12.68	1.61
2RGP	SGE	218.33	41.28	6.76	1.40
	FMS	225.70	46.36	11.77	1.77
	FMS+SGE	517.05	67.51	9.90	1.59
2VT4	SGE	166.11	36.47	7.51	1.50
	FMS	223.36	65.93	10.45	1.73
	FMS+SGE	376.70	78.04	11.07	1.67
2GTK	SGE	53.17	22.27	7.21	1.59
	FMS	613.72	75.99	10.30	1.67
	FMS+SGE	319.87	60.89	10.75	1.69
1BCD	SGE	6.40	14.38	6.35	1.67
	FMS	147.94	25.25	5.43	1.65
	FMS+SGE	43.48	25.22	7.74	1.74
1UYG	SGE	0.00	3.56	0.65	0.92
	FMS	590.68	90.43	9.32	1.62
	FMS+SGE	75.01	35.68	7.90	1.25
1L2S	SGE	0.00	0.00	0.54	1.07
	FMS	264.83	55.17	7.82	1.61
	FMS+SGE	235.40	56.64	8.84	1.48
2HZI	SGE	86.92	23.91	5.03	1.40
	FMS	149.30	23.72	4.43	1.53
	FMS+SGE	103.28	28.80	5.51	1.50
1KVO	SGE	62.81	35.80	5.84	1.29
	FMS	39.26	6.16	4.04	1.53
	FMS+SGE	51.04	32.23	8.04	1.47
1R9O	SGE	31.34	4.59	1.53	1.07
	FMS	0.00	2.17	2.45	1.20
	FMS+SGE	15.67	5.33	1.63	1.07
1E66	SGE	247.06	52.18	8.20	1.37
	FMS	11.28	3.43	1.56	1.19
	FMS+SGE	508.10	70.25	8.99	1.43
2AA2	SGE	4.13	5.00	0.74	0.45
	FMS	190.19	26.71	3.32	1.17
	FMS+SGE	62.02	7.07	0.90	0.64
1C8K	SGE	0.00	0.00	0.63	0.98
	FMS	0.00	0.00	0.47	0.83
	FMS+SGE	0.00	0.00	0.82	1.00

[a]PDB codes used with accompanying DUD-E libraries (actives + decoys). [b]FE = AUC_{curve}/AUC_{random} thus baseline random selection always yields a FE = 1.00.

contributing factor as to why FMS scoring yields enhanced enrichment involves the fact that use of a crystallographic reference captures elements of what is important for activity for at least one active ligand. Because rank-ordering of "actives" using FMS scoring are biased toward the known binder, higher enrichments can be obtained. With the addition of the SGE term, sampling and rank-ordering using FMS+SGE will change as a result of, for example MW bias, which leads to different enrichment results (less-favorable in most cases for the present tests). Overall, the enrichment tests validate the ability of FMS and FMS+SGE protocols to enrich for true actives relative to SGE alone by prioritizing molecules with similar binding profiles as a known ligand. This strongly suggests use of a pharmacophore reference to help guide virtual screening and is a viable alternative to the standard DOCK protocol.

As an additional point, in general, good enrichment should depend only on actives being ranked earlier than decoys without regards to there being "similarity" among groups of compounds. However, use of the FMS function might be expected to yield higher early similarity, compared to the entire set of actives as a whole, provided the composition of active molecules in a given database does contain subsets with 2D similarity, and a larger than average number of docked compounds yield good 3D overlap with the reference pharmacophore. To explore this issue, among rank-ordered active compounds, we computed all possible pairwise Tanimoto coefficients using the DOCK fingerprinting method motivated by the MOLPRINT algorithm[43,44] and plotted the data as heatmaps (Figure 17).

While additional studies should be pursued, especially those employing more than one reference per system as was done in the current study, Figure 17 reveals that in a number of cases, active molecules do in fact appear to have higher similarity earlier in rank-ordered list when using FMS versus SGE scoring (Figure 17, red/yellow vs blue, top vs bottom rows). Rank-ordering with FMS also shows a tendency to cluster similar molecules together. Particularly interesting examples include 1SJO, 1UYG, and 1L2S for which SGE shows poor (random in two cases) enrichment compared to FMS as gauged by the shape of the ROC curves in Figure 16.

4.5. Case Studies Targeting EGFR, IGF-1R, and HIVgp41. To further gauge the utility of using FMS methods, we rescored virtual screening results for three systems being targeted in our laboratory: epidermal growth factor (EGFR),[45,46] insulin-like growth factor 1 receptor (IGF-1R), and human immunodeficiency virus glycoprotein 41 (HIVgp41)[9,47] and visually examined the number of pharmacophore matches for top-ranked molecules under different conditions (Figure 18). The FMS references employed for EGFR (erlotinib) and IGF-1R (isoquinolinedione analog) were based on known small molecule inhibitors, while the HIVgp41 reference was based on four key amino acid side chains (WWDI) from a known peptide inhibitor. The receptors and accompanying references were derived from crystallographic structures (PDB codes 1M17, 2ZM3, and 1AIK, respectively), and the molecules docked to each target were taken from the publically available ZINC[48] collection of purchasable organic compounds. For each screen, the top 100000 ranked compounds obtained with the standard docking protocol (grid score with FLX protocol) were retained and then rescored and reranked using DOCK Cartesian energy (DCE, which is comparable to SGE but in Cartesian space), FMS, and FMS+DCE scoring protocols.

As shown in Figure 18, the number of pharmacophores matched for the top 25 ranked compounds is relatively small using DCE. In sharp contrast, use of FMS or FMS+DCE show, for example, many more matched HBD (blue arrows), HBA (red arrows), ARO (orange arrows), and PHO (cyan spheres) features. It is important to note that the plots in Figure 18 show how many "matched" pharmacophores were obtained, relative

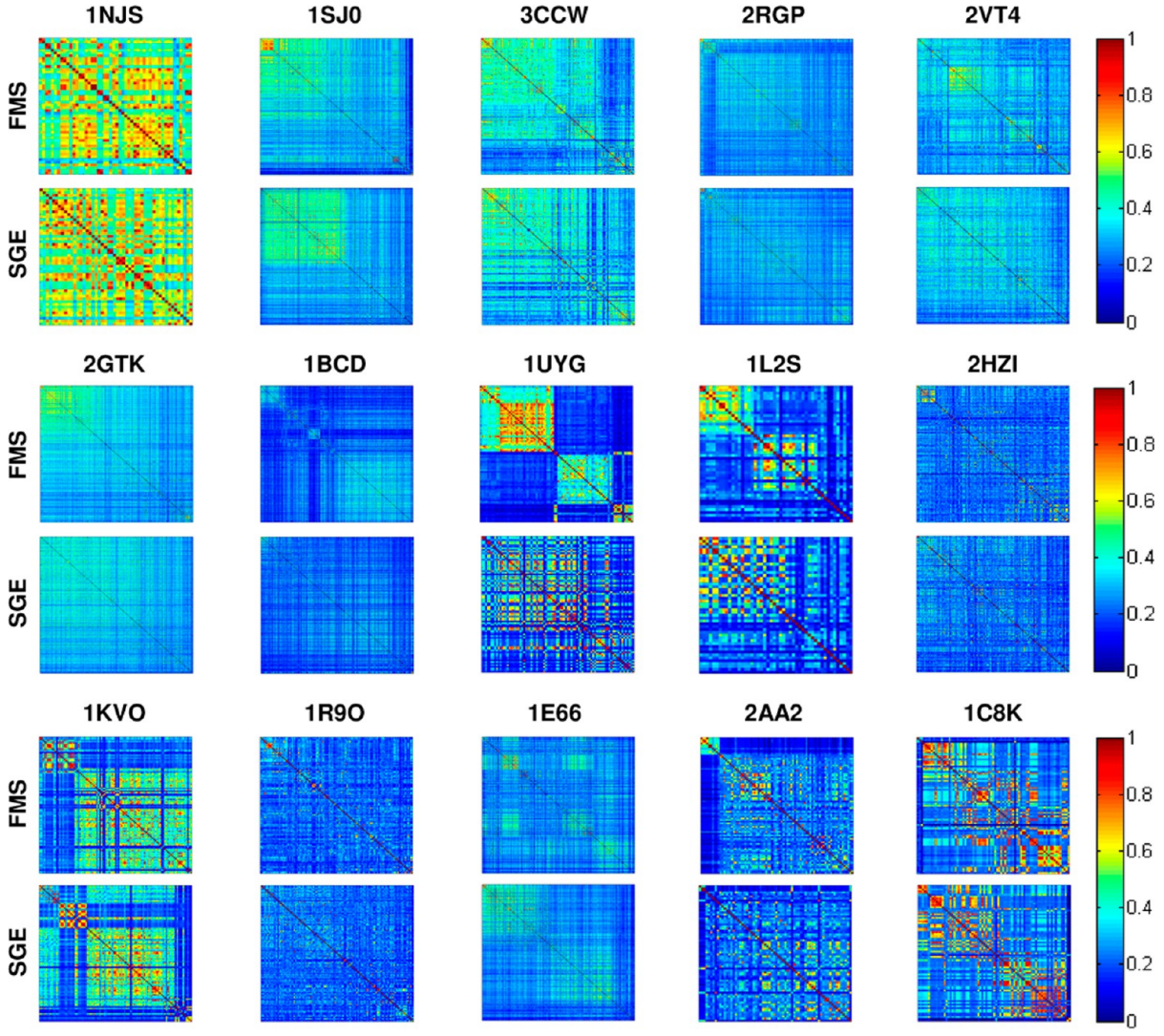


Figure 17. Pairwise Tanimoto heatmap for 15 DUD-E systems using FMS (top) and SGE (bottom) protocol. The color scheme in the heatmap represents the magnitude of Tanimoto similarity, and the *x*/*y* axis represents the rank-ordered list (FMS or SGE) of unique active molecules for each system.

to the reference, but candidate compounds can contain "unmatched features" that extend beyond the volume defined by the reference compound; the functional form of eq 1 does not necessarily penalize unmatched features relative to the candidate. This behavior could be changed, for example, by including simply a penalty term based on the number of unmatched groups in the candidate; however, this was not explored in great detail. Other functional forms besides eq 1 could also be investigated. In any event, the number of matched and unmatched features, including types, for each docked pose, are printed to the DOCK output, which can be useful to determine whether particular characteristics have been satisfied.

As a specific example, an interesting result from the present analysis is a lack of matched pharmacophore features to the Asp carboxylate group in the HIVgp41 reference (Figure 18 row C). An examination of ranked poses higher up the FMS and FMS +DCE lists did indeed reveal compounds with overlap to the reference carboxylate but they were not ranked as well as compounds with multiple matches involving two Trp indole rings and a hydrophobic Ile (Figure 18, row C). Given the biological importance of the Asp group in this system, an effective small molecule mimic would reasonably be expected to contain a negatively charged or hydrogen-bonding group at this position.[49,50] A straightforward way to enforce this requirement was devised by using a modified HIVgp41 reference that simply included 5 copies of the Asp carboxylate which had the effect of weighting this feature more heavily, as shown in Figure 18, row D. For this particular test, weighting the Asp more highly had the desired effect but at the expense of losing hydrophobic matches to Ile (Figure 18, FMS and FMS+DCE, row C vs D). As a general point, this example demonstrates the ease with which specific pharmacophore features can be emphasized over others using the current DOCK infrastructure.

Finally, in terms of additional ligand properties, Figure 19 plots results from the HIVgp41 screen for different groups of top-ranked molecules (N = 500) each obtained by one of the ranking protocols. Consistent with previous studies from our laboratory, use of DCE (or SGE) shows a bias toward larger molecules. In contrast, compounds ranked by FMS score are smaller in size as demonstrated by ligands with lower molecular weights (Figure 19d) and fewer numbers of rotatable bonds (Figure 19e). As anticipated, use of FMS+DCE yields molecular weights and numbers of rotatable bonds roughly in-between DCE and FMS. For scoring, use of DCE results in

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

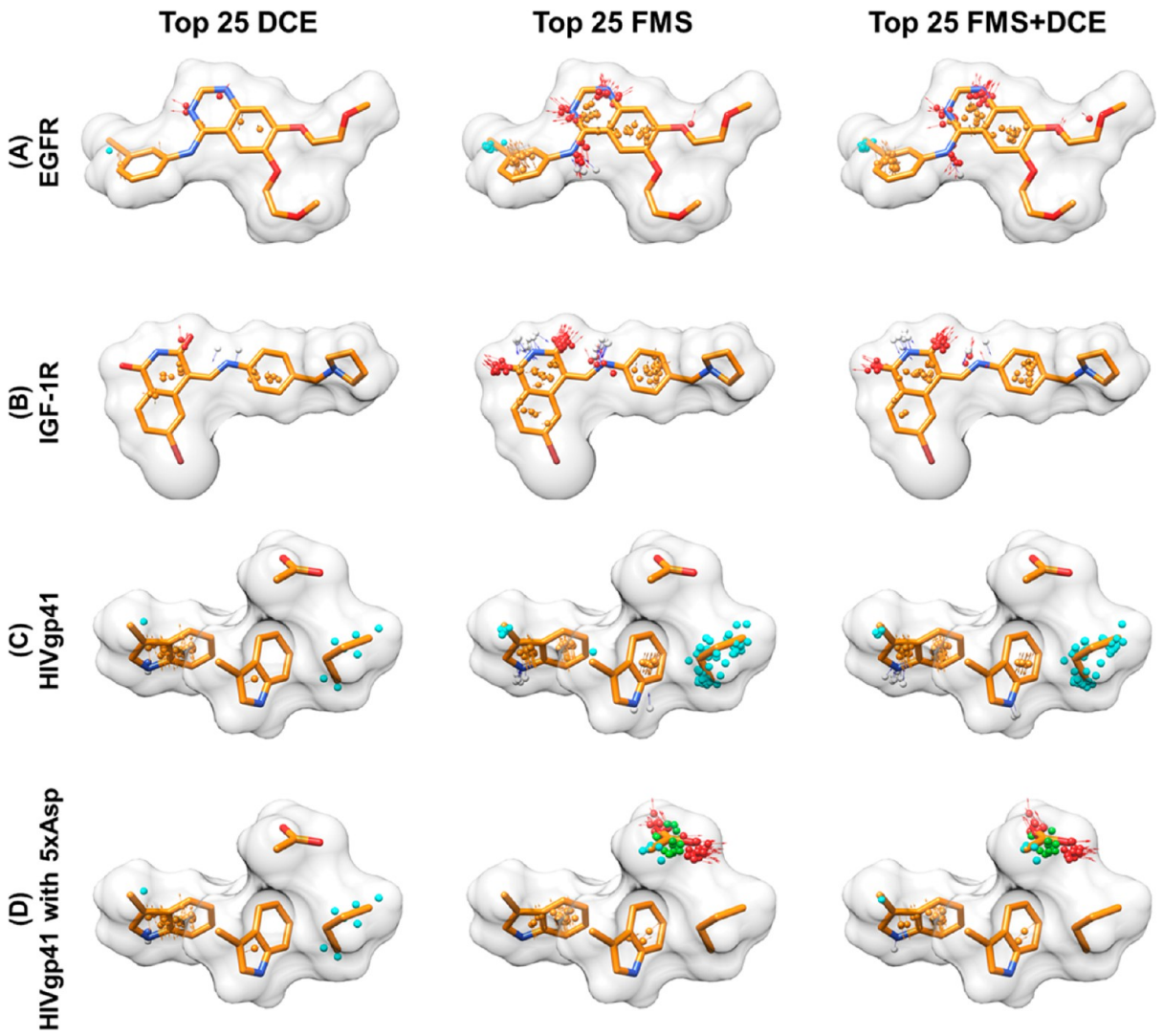


Figure 18. References (orange sticks, gray surface) used to rescore virtual screening results targeting (A) EGFR, (B) IGF-1R, (C) HIVgp41, and (D) HIVgp41 with Asp side chain weighted 5 times. Matched pharmacophore features include: PHO in cyan, HBA (vertex and vector) in red, HBD vector in blue, hydrogen vertex in gray, ARO (vertex and vector) in orange, POS in magenta, and NEG in green (see Theoretical Methods for definitions).

more favorable DCE energies (Figure 19a, blue vs red or green), FMS results in more favorable FMS scores (Figure 19b, red vs blue or green), and FMS+DCE results in more favorable FMS+DCE scores (Figure 19c, green vs blue or red). And, rescoring molecules obtained with one function with another function leads to the expected results. For example, DCE score distributions for top-ranked FMS+DCE molecules are in between that of DCE and FMS (Figure 19a green), FMS score distributions for top-ranked FMS+DCE molecules are in between that of FMS and DCE (Figure 19b green), and FMS +DCE score distributions for top-ranked FMS molecules are in between that of FMS+DCE and DCE (Figure 19c red). Importantly, use of the combined FMS+DCE function to rescore virtual screening results yield both favorable FMS scores and DOCK energies. This suggests use of a reference to rescore screening results could also be a viable way to identify compounds that make known interaction patterns, with favorable interaction energies, while reducing molecular weight bias.

5. CONCLUSION

In conclusion, the primary goal of this study was to develop, implement, and thoroughly test a pharmacophore-based scoring function for the docking program DOCK. The resulting method, termed pharmacophore matching similarity (FMS) score, was validated using experiments that help gauge accuracy relative to the standard DOCK single-energy grid (SGE) protocol, and the combination score FMS+SGE. Three groups of validation experiments were performed: (i) pose reproduction (Figures 7−11 and Table 4), (ii) crossdocking (Figures 12−15), and (iii) enrichment (Figure 16 and Table 5). Importantly, in terms of pose reproduction, use of FMS (93.5%) or FMS+SGE (98.3%) functions yielded significantly higher success rates than the standard SGE (72.5%) method when evaluated using 1043 systems in the SB2012 testset. The nearly perfect success rate obtained with the combined FMS +SGE function, which biases sampling to match a reference while simultaneously including energetic constraints imposed by a binding site, is notable and strongly suggests the method will have applicability for structure-based drug design provided a "suitable" reference can be identified. Tests using FMS alone

dx.doi.org/10.1021/jp506555w | J. Phys. Chem. B 2015, 119, 1083−1102

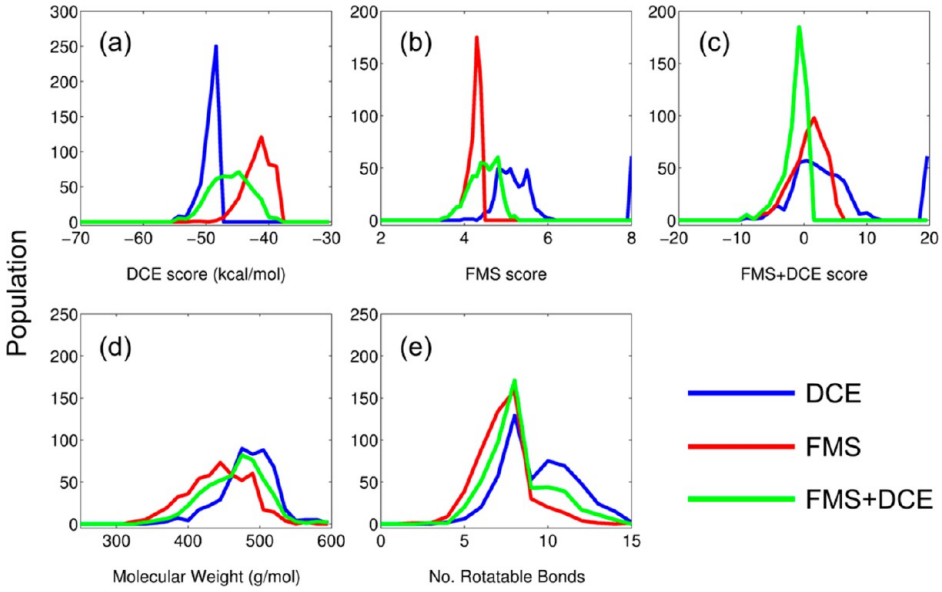


Figure 19. Histograms of rescoring results for the top 500 molecules selected from virtual screening targeting HIVgp41.

for pose reproduction showed relatively few ligand poses falling into false positive (FP) and false negative (FN) regions defined by quadrant partition using specific RMSD and FMS score cutoff criteria (Figure 9). Interestingly, visual examination of the worst FP cases (Figure 10) revealed, in most instances, that the candidate and references poses were in fact well-overlaid and that only one part of the molecule was not well-matched. Unlike the standard DOCK energy function, the geometry-based FMS scores show reasonable correlation with RMSD.

For crossdocking, while use of FMS scoring alone showed significant improvement with regards to systems on the diagonal (cognate protein−ligand pairs), the overall matrix success rate in 4 out of 6 cases was significantly lower than SGE. Examination of the underlying reference structures showed that FMS docking success is highly dependent on how well the pharmacophore reference overlays with the RMSD reference (Figures 12−15). Thus, while use of FMS scoring alone to drive sampling of a ligand using a reference without possibilities for good overlap yields poor results, such behavior makes physical sense. More importantly, the results dramatically emphasize that the FMS function works best when the goal is identification of molecules that resemble the reference, as was the original intent. As expected, use of the combined FMS+SGE function provides more of a balance and yields the highest crossdocking matrix success rates (Figure 12).

In terms of enrichment, receiver operator characteristic (ROC), area under the curve (AUC), and fold enrichment (FE) analyses, in general, showed that FMS and FMS+SGE functions yield better performance than SGE alone (and random selection) for both early and total enrichment (Figure 16 and Table 5) when evaluated over 15 systems taken from the DUD-E database. For several systems, FMS+SGE enrichment appears roughly in between that obtained using FMS or SGE alone (Figure 16). Importantly, FE values computed very early in rank-ordered lists (0.1% and 1%) showed using FMS and FMS+SGE yielded 10−13 out of 15 FE values enhanced relative to the standard protocol SGE (Table 5, columns A and B) despite the fact that only a "single" reference (cognate ligand) was used to guide sampling of compounds. Future studies should evaluate enrichment outcomes using multiple FMS references.

In terms of virtual screening, rescoring results obtained from standard docking to three target of pharmaceutical interest (EGFR, IGF-1R, and HIVgp41) showed that the FMS and FMS+DCE (equivalent to FMS+SGE) methods yielded more compounds with greater numbers of pharmacophore matches when the top 25 compounds from each method were examined (Figure 18). The example also demonstrated how FMS scoring could utilize small organic molecules or noncontiguous protein side chains as a reference. For gp41 in particular, examination of top poses revealed that none of the compounds matched an important Asp side chain in the initial pharmacophore model. A simple modification of the reference to include multiple copies of the Asp weighted this functionality more highly, and when rescored, yielded top-ranked compounds with the desired interaction. Importantly, this result further establishes the importance of the FMS "reference" in addition to demonstrating how pharmacophores could be customized.

Finally, the current results suggest several directions for future research, including exploring other functional forms of the main FMS equation (eq 1), testing FMS score in combination with other scoring functions (i.e., footprint similar scoring), development of a receptor-based[51] as opposed to the current ligand-based method, and implementation of routines to address multiple pharmacophore references simultaneously.[52] Ongoing work involves incorporation of FMS scoring into a *de novo* design version of DOCK, currently under development in our laboratory, to allow pharmacophore-guided *de novo* growth of new ligands from scratch having similar binding profiles as a known reference.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rizzorc@gmail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Gratitude is expressed to Trent E. Balius for help in early code development and computational assistance, to Yulin Huang for computational assistance, and to William J. Allen for helpful discussions and critical reading of the manuscript. The research utilized resources at the New York Center for Computational Sciences at Stony Brook University/Brookhaven National Laboratory, which is supported by the U.S. Department of Energy under Contract DE-AC02-98CH10886 and by the State of New York. The work was supported by the National Institutes of Health, Grant R01GM083669.

■ REFERENCES

(1) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining Techniques to Model RNA-Small Molecule Complexes. *RNA* **2009**, *15*, 1219−1230.

(2) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and Validation of a Modular, Extensible Docking Program: Dock 5. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601−619.

(3) Klebe, G. Virtual Ligand Screening: Strategies, Perspectives and Limitations. *Drug Discovery Today* **2006**, *11*, 580−594.

(4) Kuntz, I. D. Structure-based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078−1082.

(5) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.

(6) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862−865.

(7) Balius, T. E.; Mukherjee, S.; Rizzo, R. C. Implementation and Evaluation of a Docking-rescoring Method Using Molecular Footprint Comparisons. *J. Comput. Chem.* **2011**, *32*, 2273−2289.

(8) Balius, T. E.; Allen, W. J.; Mukherjee, S.; Rizzo, R. C. Grid-Based Molecular Footprint Comparison Method for Docking and De Novo Design: Application to HIVgp41. *J. Comput. Chem.* **2013**, *34*, 1226−1240.

(9) Holden, P. M.; Kaur, H.; Goyal, R.; Gochin, M.; Rizzo, R. C. Footprint-Based Identification of Viral Entry Inhibitors Targeting HIVgp41. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3011−3016.

(10) Berger, W. T.; Ralph, B. P.; Kaczocha, M.; Sun, J.; Balius, T. E.; Rizzo, R. C.; Haj-Dahmane, S.; Ojima, I.; Deutsch, D. G. Targeting Fatty Acid Binding Protein (FABP) Anandamide Transporters: A Novel Strategy for Development of Anti-Inflammatory and Anti-Nociceptive Drugs. *PLoS One* **2012**, *7*, e50968.

(11) Ehrilich, P. Uber die Constitution des Diphtheriegiftes. *Deut. Med. Wochenschr.* **1898**, *24*, 597−600.

(12) Ehrilich, P. Uber den Jetzigen Stand der Chemotherapie. *Ber. Dtsch. Chem. Ges.* **1909**, *42*, 17−47.

(13) Güner, O. F.; Bowen, J. P. Setting the Record Straight: The Origin of the Pharmacophore Concept. *J. Chem. Inf. Model.* **2014**, *54*, 1269−1283.

(14) Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations, 1998). *Pure Appl. Chem.* **1998**, *70*, 1129−1143.

(15) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.* **2010**, *53*, 539−558.

(16) Yang, S. Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discovery Today* **2010**, *15*, 444−450.

(17) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. A. F.; Klomp, J. P. G.; de Vlieg, J.; Del Rio, A. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, *52*, 1607−1620.

(18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(19) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of Common Functional Configurations among Molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563−571.

(20) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532−549.

(21) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45*, 160−169.

(22) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647−671.

(23) Richmond, N. J.; Abrams, C. A.; Wolohan, P. R.; Abrahamian, E.; Willett, P.; Clark, R. D. GALAHAD: 1. Pharmacophore Identification by Hypermolecular Alignment of Ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567−587.

(24) Joseph-McCarthy, D.; Alvarez, J. C. Automated Generation of MCSS-Derived Pharmacophoric DOCK Site Points for Searching Multiconformation Databases. *Proteins: Struct. Funct. Bioinf.* **2003**, *51*, 189−202.

(25) Joseph-McCarthy, D.; Thomas, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based Molecular Docking to Account for Ligand Flexibility. *Proteins: Struct. Funct. Bioinf.* **2003**, *51*, 172−188.

(26) *MOE*; version 2012.10, Chemical Computing Group Inc.: Montreal, Canada, 2012.

(27) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986−2000.

(28) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.

(29) Shoichet, B. K.; Kuntz, I. D. Matching Chemistry and Shape in Molecular Docking. *Protein Eng.* **1993**, *6*, 723−732.

(30) *Tripos*, Mol2 file format; Certara: St. Louis, MO, 2009.

(31) *DOCK6.6*, user manual; http://dock.compbio.ucsf.edu (accessed Sept 23, 2014).

(32) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera: A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(33) ACD/ChemSketch, version 14.01; Advanced Chemistry Development, Inc.: Toronto, ON, Canada, 2014 (http://www.acdlabs.com).

(34) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(35) Bunke, H.; Shearer, K. A Graph Distance Metric Based on the Maximal Common Subgraph. *Pattern Recognition Letters* **1998**, *19*, 255−259.

(36) Bunke, H. Graph Matching: Theoretical Foundations, Algorithms, and Applications. *International Conference on Vision Interface* **2000**, 82−84.

(37) Allen, W. J.; Rizzo, R. C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-based Design. *J. Chem. Inf. Model.* **2014**, *54*, 518−529.

(38) *DMS*; UCSF Computer Graphics Laboratory: San Francisco, CA, http://www.cgl.ucsf.edu/Overview/software.html (accessed May 4, 2010).

(39) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722−729.

(40) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(41) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(42) Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a Pose Generation and Database Enrichment Tool. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 749−773.
(43) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708−1718.
(44) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* **2003**, *44*, 170−178.
(45) Balius, T. E.; Rizzo, R. C. Quantitative Prediction of Fold Resistance for Inhibitors of EGFR. *Biochemistry* **2009**, *48*, 8435−8448.
(46) Huang, Y.; Rizzo, R. C. A Water-Based Mechanism of Specificity and Resistance for Lapatinib with ErbB Family Kinases. *Biochemistry* **2012**, *51*, 2390−2406.
(47) Holden, P. M.; Allen, W. J.; Gochin, M.; Rizzo, R. C. Strategies for Lead Discovery: Application of Footprint Similarity Targeting HIVgp41. *Bioorg. Med. Chem.* **2014**, *22*, 651−661.
(48) Irwin, J. J.; Shoichet, B. K. ZINC: A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2004**, *45*, 177−182.
(49) Allen, W. J.; Rizzo, R. C. Computer-Aided Approaches for Targeting HIVgp41. *Biology* **2012**, *1*, 311−338.
(50) He, Y.; Liu, S.; Li, J.; Lu, H.; Qi, Z.; Liu, Z.; Debnath, A. K.; Jiang, S. Conserved Salt Bridge Between the N- and C-terminal Heptad Repeat Regions of the Human Immunodeficiency Virus Type 1 gp41 Core Structure is Critical for Virus Entry and Inhibition. *J. Virol.* **2008**, *82*, 11129−11139.
(51) Hu, B.; Lill, M. A. Exploring the Potential of Protein-based Pharmacophore Models in Ligand Pose Prediction and Ranking. *J. Chem. Inf. Model.* **2013**, *53*, 1179−1190.
(52) Gardiner, E. J.; Cosgrove, D. A.; Taylor, R.; Gillet, V. J. Multiobjective Optimization of Pharmacophore Hypotheses: Bias Toward Low-energy Conformations. *J. Chem. Inf. Model.* **2009**, *49*, 2761−2773.