Receptor Flexibility in the in Silico Screening of Reagents in the S1′ Pocket of Human Collagenase

Per Källblad,* Nikolay P. Todorov, Henriëtte M. G. Willems, and Ian L. Alberts

De Novo Pharmaceuticals Ltd., Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR, U.K.

Received October 13, 2003

A major difficulty in structure-based molecular design is the prediction of the structure of the protein−ligand complex because of the enormous number of degrees of freedom. Commonly, the target protein is kept rigid in a single low-energy conformation. However, this does not reflect the dynamic nature of protein structures. In this work, we investigate the influence of receptor flexibility in virtual screening of reagents on a common scaffold in the S1′ pocket of human collagenase (matrix metalloproteinase-1). We compare screening using a single-crystal structure and multiple NMR structures, both apo and holo forms. We also investigate two computational methods of addressing receptor flexibility that can be used when NMR data are not available. The results from virtual screening using the experimental structures are compared to those obtained using the two computational methods. From the results, we draw conclusions about the impact of target flexibility on the identification of active and diverse reagents in a virtual screening protocol.

Introduction

Molecular docking in a virtual screening program seeks to explore the binding mode and affinity of a large number of ligands from commercial or in-house databases. It is well-known that ligands can induce significant changes in protein structure upon binding,[1,2] and changes in just a few receptor side chain conformations can lead to false negatives in the screening process. Ignoring flexibility will inevitably reduce the number and diversity of successfully screened molecules. Thus, to model correctly the binding characteristics of known inhibitors and to design novel leads with different binding characteristics, it is crucial to incorporate receptor flexibility into the drug design process.[3]

Two distinct methods have been used to deal with receptor flexibility in the area of ligand docking. First, an ensemble of preselected receptor models can be used.[4,5] Multiple protein conformations have been used in recent studies for binding mode prediction[6] and pharmacophore design.[7] Procedures have also been developed for identifying flexible protein substructures for use in ligand docking and database screening.[8,9] Second, the receptor has been allowed to flex partially during the docking procedure. This has been limited to specific degrees of freedom of the receptor, such as hydrogen atoms, hydrogen bond donor/acceptor groups, and side chain rotamers.[10] Alternatively, full relaxation of the complex by minimization can be conducted[11] starting from a number of initial positions for the ligand. Protein conformational flexibility is clearly an important aspect of ligand binding;[12] however, it is not yet a standard of the ligand-docking protocol.

We investigate the influence of receptor flexibility in virtual screening. We compare screening using a single-crystal structure and multiple NMR structures, both apo and holo forms. We also investigate two computational methods of addressing receptor flexibility: (i) a preselection of receptor conformations from an exhaustive exploration of side chain flexibility; (ii) a dynamic exploration of side chain movements of the known crystal structure. The virtual screening results using the experimental structures are compared with those obtained using the two computational modeling methods. From the results we illustrate the importance of receptor flexibility in identifying additional compounds.

These methods are applied in a virtual screen of reagents to a common scaffold in the S1′ pocket of human collagenase (matrix metalloproteinase-1 (MMP-1)). MMPs have been the subject of intense study because of their involvement in a number of pathogenic states,[13] including cancer[14] and arteriosclerosis.[15] Potent MMP inhibitors have been developed including peptidic inhibitors comprising hydroxamate, carboxylate,[16] and phosphinic[17] zinc-chelating groups. Significant conformational flexibility has been ascribed to the S1′ pocket, and it is known to undergo conformational changes in multiple side chains upon binding of certain ligands.[17−21] The S1′ pocket has also proved to be of key importance for the binding of endogenous protein inhibitors.[22]

Methods

Input Structure Set. The MMP-1 models used in the virtual screen are listed in Table 1. The experimental structures included one template crystal structure (X-ray) and 30 conformers each from an unbound (NMR_{Apo}) and a bound NMR model (NMR_{Holo}). The following six residues that can affect the shape and ligand-binding properties of the S1′ pocket were allowed conformational flexibility in our investigations: Leu181, Arg214, Val215, Ser239, Tyr240, and Phe242 (numbering as in SWISS-PROT[26]). All water molecules, synthetic ligands, and non-Zn metal atoms were removed. Hydrogen atoms were added to all the structures in InsightII/Discover3 (Accelrys Inc., San Diego, CA) at pH 7.0. Their positions were subject to 500 iterations of steepest descent minimization followed by 500 steps of conjugate gradient minimization using the CFF

* To whom correspondence should be addressed. E-mail: Per.Kallblad@denovopharma.com. Phone: +44 1223 238000. Fax: +44 1223 238088.

10.1021/jm031061l CCC: $27.50
Published on Web 04/28/2004

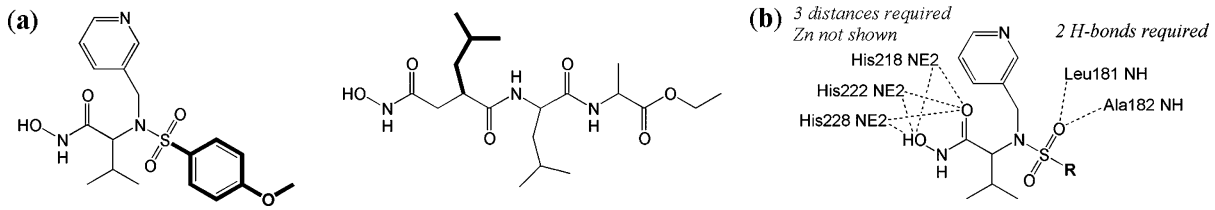

Figure 1. (a) Ligands removed from MMP-1 models. Bold parts occupy the S1′ pocket. Shown on the left is CGS-27023A (IC_{50} = 27 nM) from NMR_{Holo} (4AYK). Shown on the right is RO 31-4724 (IC_{50} = 40 nM) from X-ray (2TCL). (b) Scaffold and distance requirements (experimentally observed distances and H-bond mimics) used in the virtual screen.

Table 1. MMP-1 S1′ Models Used in Virtual Screen

model	resolution	*n*	S1′ ligand	PDB code	template	ref
X-ray	2.2	1	RO 31-4724	2TCL	n/a	23
NMR_{Apo}	n/a	30		1AYK	n/a	24
NMR_{Holo}	n/a	30	CGS-27023A	4AYK	n/a	19
Dynasite	n/a	15		n/a	2TCL	25
Reflex	n/a	n/a		n/a	2TCL	n/a

force field.[27] All non-hydrogen atoms were held rigid during these minimizations.

Side Chain Flexibility. The backbone-dependent rotamer library of Dunbrack and co-workers[28,29] was used in both flexible receptor methods. Rotamers with backbone $\phi-\psi$ angles within 5° of the template structure were chosen from the library, and only those that did not clash with any of the rigid backbone atoms or with any nonflexible side chain atom were retained. The van der Waals radii were assigned from the Bondi set,[30] and a heavy atom van der Waals overlap of 0.7 Å was allowed. This led to 6, 11, 3, 3, 3, and 4 possible rotamers for Leu181, Arg214, Val215, Ser239, Tyr240, and Phe242, respectively.

Site Ensemble. The program Dynasite[25] was developed to sample efficiently local side chain flexibility. The procedure for the ensemble generation followed that of Källblad and Dean,[25] and only a brief summary of the procedure will be given here. Out of 7128 theoretical solutions, an exhaustive ensemble of 2115 nonclashing conformers was generated with alternative rotameric combinations for the flexible residues. The six flexible side chains plus the side chains of nine adjacent residues (Glu201, Trp203, Phe207, Asn211, Leu235, Tyr237, Thr241, Ser243, Val246) were minimized while all other atoms were held rigid. A subset of the 2115 minimized conformers was selected through a combination of principal component analysis (PCA) and clustering. The final Dynasite ensemble consisted of 15 structurally diverse conformers that were low in energy and representative of the minimized ensemble. The Dynasite ensemble has been validated against experimentally determined structures.[25]

Geometric Variation within and between Ensembles. PCA[31] was used to examine the most significant trends of the geometric variation between the template crystal structure (n = 1), the NMR models (n = 30 + 30), and the Dynasite ensemble (n = 15). The input data matrix (76 × 76) consisted of pairwise root mean square deviation (rmsd) values over the side chain atoms of the flexible residues. Principal components with eigenvalues greater than 1.0 were considered to be significant.

Virtual Screen. Five different target sets were employed: (i) the static X-ray structure; (ii) the 30 NMR_{Apo} models; (iii) the 30 NMR_{Holo} models; (iv) the 15 conformers of the Dynasite ensemble; (v) the flexible crystal structure (Reflex; see below). See Table 1 for more details. The ligands from the X-ray model and the NMR_{Holo} models were removed (Figure 1a). In all cases, the virtual screen involved reagents attached to a common scaffold.

The scaffold (*N*-sulfonylamino acid hydroxamate[32]) was extracted from the NMR_{Holo} structure. It was allowed flexibility in the site during the virtual screen but was anchored through experimentally observed distances and significant H-bonding interactions with the site (Figure 1b). The substituent R was assumed to be derived from a sulfonyl chloride. Some sulfonyl chlorides were therefore extracted from the Daylight Available Chemicals Directory (ACD) (Daylight Chemical Information Systems Inc., CA). Sulfonyl chlorides containing atoms other than C, N, O, S, Na, Ca, K, or halogens were removed from the set, as were molecules containing chains of more than five sequential carbon atoms, molecules with two sulfonyl chloride groups, acid chlorides, azides, aldehydes, and hydrazones, and molecules with MW > 350. A further 16 sulfonyl chlorides used to synthesize known MMP-1 inhibitors were added to form the final set of 254 reagents. For each reagent, the physiologically relevant protonation states were selected. Ring conformations within 15 kJ/mol of the global minimum were explored in the screening. The reagents were attached to the scaffold at substitution position R depicted in Figure 1b.

The de novo drug design program Skelgen was used for the virtual screen in all five target sets. A brief discussion of the Skelgen algorithm is given below and further details of the principles involved as well as practical applications of the method can be found in previous publications.[33,34]

Skelgen takes as input the three-dimensional coordinates of the receptor, a rectangular box that defines the binding site pocket and a predefined set of fragments. The algorithm invokes a simulated annealing procedure to minimize a function that assesses the suitability of molecular structures within the active site box. In this process, a set of initial fragments are randomly selected and connected together. The putative ligand is then altered by several transitions, which include rigid-body displacements, conformational changes, and fragment addition, deletion, and replacement. The transition type is randomly chosen, and at each step the value of the objective function is determined. The modified structure is accepted or rejected on the basis of the Metropolis condition, $p = \exp(\Delta F/T)$, where ΔF represents the change in value of the objective function resulting from the transition, T is the temperature, which is lowered during the annealing procedure, and p is the probability of acceptance of the transition. The length of each Skelgen annealing run is defined by the number of transitions at each temperature setting, termed the Markov chain length, and by the number of Markov chains as the temperature is depressed. The algorithm generates one structure for each run, and as a result of the stochastic nature of the simulation, different random starts can lead to different solutions.

Skelgen was extended to incorporate dynamic flexibility in the reagent screening process by allowing an additional transition type in the simulated annealing protocol corresponding to side chain mobility. In this approach, the nonclashing rotamers for each flexible side chain as detailed above are retained and the space of the 2115 receptor conformations is sampled during each annealing run (data not shown). Initially, a rotamer is chosen at random for each of the flexible residues and, during the annealing procedure, if the new side chain rotation move is picked, one of the six flexible residues

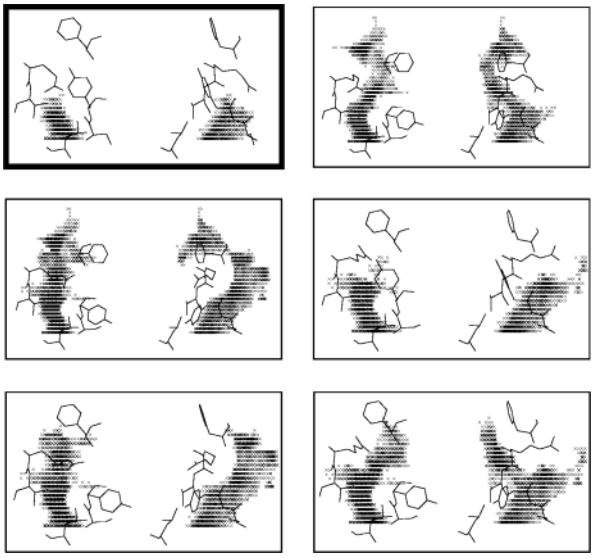

Figure 2. Five sample cavities of the Dynasite S1′ ensemble. The right view is rotated 90° around the *Y* axis. The cavity of the template X-ray structure (which also was reproduced by Dynasite) has a bold frame.

is randomly chosen and a new rotamer is randomly assigned. As for the other transition types, the move is scored and either accepted or rejected on the basis of the simulated annealing protocol. The length of each Markov chain was doubled when protein as well as ligand flexibility was invoked; however, the annealing cooling temperature schedule was unchanged. The name Reflex denotes the use of Skelgen with the receptor flexibility switched on.

In this work, Skelgen was used in reagent screening mode such that each reagent in our data set corresponding to a fragment was individually subject to a Skelgen simulated annealing run, and thus, transition types for fragment addition, deletion, and replacement were switched off. Only transition types for ligand conformational changes were retained. The objective function comprises several terms, which are conferred a zero value when the appropriate constraints are satisfied and conferred a positive value otherwise. The constraints used in this work included intramolecular ligand van der Waals clashes (according to the Bondi set[30]), intermolecular ligand−receptor van der Waals clashes (also according to the Bondi set[30]), and experimentally observed distances and H-bond mimics (2.5−3.5 Å) between the scaffold and site (see Figure 1b). The transition types during the simulated annealing process are ligand conformational moves for the X-ray, $NMR_{Apo/Holo}$, and Dynasite ensemble targets, and additionally, side chain moves are included for the flexible crystal target (Reflex). In the latter case, rotamer moves are given the same selection frequency weighting as the ligand conformational changes. As an example of this protocol, in flexible receptor mode, if a rotamer is selected that clashes with the reagent, an intermolecular ligand−receptor van der Waals clash will be added to the score. The move, in this case a rotamer change, will either be accepted or rejected on the basis of the Metropolis condition. Each reagent was screened 10 times to ensure proper sampling of the stochastic process. A reagent was deemed to be successful in the screen, yielding a "hit", if it satisfied the constraints in at least 1 of the 10 runs.

Results

Geometric Variation within and between Ensembles. The size and shape of the S1′ pocket have been established as the key to ligand binding to the MMPs. Figure 2 shows sample cavity shapes of five members of the Dynasite ensemble in comparison to that of the template crystal structure. It can be seen that the template crystal structure (X-ray) has a small cavity (42 Å^3). However, the Dynasite ensemble clearly exhibits significant variation in the S1′ cavity shapes with cavity volumes ranging from 43 to 148 Å^3. The receptor conformations derived from the Reflex algorithm showed cavity diversity similar to that of the Dynasite ensemble (data not shown).

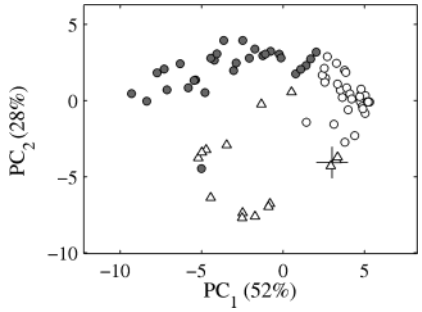

Figure 3. Object distribution along the two most significant components (PC_{1-2}) of the PCA model of the NMR and Dynasite ensemble. The amount of variation explained by each PC is indicated. Symbol key is the following: cross = X-ray; triangles = Dynasite ensemble; circles = NMR_{Apo}; filled circles = NMR_{Holo}.

The PCA model for the 76 conformers (the template X-ray structure, 30 from each NMR model and 15 from the Dynasite ensemble) consisted of four significant components explaining 91.8% of the total geometric variation in the site residues. The distribution of the conformers along the two most significant components is shown in Figure 3. Along PC_1 there is a clear separation between the two NMR models reflecting the introduction of the ligand. The reference crystal structure, 2TCL, denoted by a cross in Figure 3, overlaps with a set of distinct Dynasite conformers along PC_{1-2}. The Dynasite ensemble also spans a considerable part of the space overlapping with both NMR_{Apo} and NMR_{Holo}. The clearest separation between the Dynasite ensemble and the NMR models appears along PC_2. The conformers generated from the Reflex algorithm essentially occupy the same space as the Dynasite ensemble but with more intermediate conformers (data not shown). In terms of dihedral angles, the Dynasite ensemble covered most of the diversity observed in the NMR models. In fact, of the six flexible residues, only Arg214 and Ser239 were found to be undersampled in χ_1 space for the Dynasite ensemble. Further details of the conformational diversity of the Dynasite ensemble are given in the earlier work by Källblad and Dean.[25] The NMR models also exhibit backbone movements, which may be the origin of the increased side chain diversity[35] in these structures and is expected to have a crucial impact on the reagent screening.

Virtual Screen. The number of successful reagents for the experimental models in the virtual screen was 132, 199, and 240 for the X-ray, NMR_{Apo}, and NMR_{Holo} models. Clearly, target flexibility has a significant impact on virtual hit rates with the ensembles yielding significantly higher virtual hit rates than the static crystal structure in this screening protocol. Although not identifying as many virtual hits as the NMR ensembles, the Dynasite and Reflex models outperformed the static X-ray structure with 164 and 173 virtual hits, respectively. The high virtual hit rates in the NMR models, particularly the NMR_{Holo} ensemble, are perhaps not surprising because they show the most

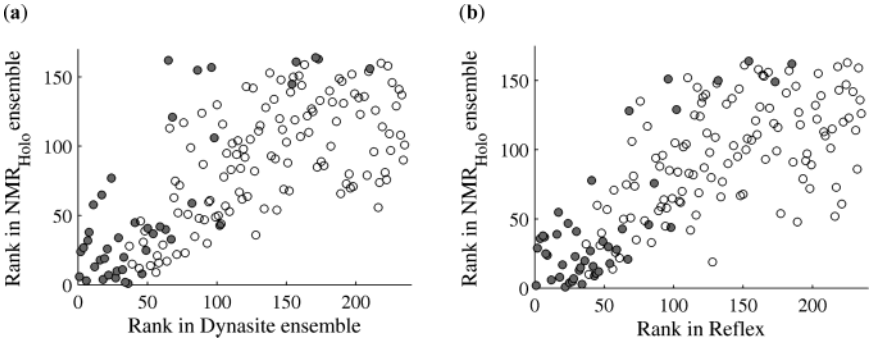

Figure 4. Correlation in reagent interaction score rank between ensembles: (a) correlation between Dynasite and NMR_{Holo} ensembles ($r^2 = 0.50$); (b) correlation between Reflex and NMR_{Holo} ensembles ($r^2 = 0.52$). Symbol key is the following: empty circle = reagent successful in X-ray, Dynasite/Reflex, and NMR_{Holo}; filled circle = reagent successful in Dynasite/Reflex and NMR_{Holo} but not in X-ray.

Table 2. Activities and Virtual Screen Results of Substituents Highlighted in Figure 5[a]

compd	WDI identifier	MMP-1 activity,	X-ray	NMR_{Apo}	NMR_{Holo}	Dynasite	Reflex
1	WD-94-007449	$K_i = 33$	8 −30.7 22	9 −28.5 52	9 −35.3 94	9 −34.8 60	9 −31.7 45
2	WD-2000-010680	$IC_{50} = 121$	9 −36.1 5	9 −29.2 45	9 −38.0 61	9 −39.1 29	9 −36.5 22
3	WD-2001-000648	$IC_{50} = 920$	7 −30.6 24	9 −30.9 29	9 −40.3 45	8 −36.4 46	9 −29.8 60
4	WD-2003-007509	$IC_{50} = 138$	8 −28.2 39	8 −28.6 50	9 −39.1 48	8 −37.3 39	8 −33.1 40
5	WD-98-013532	$IC_{50} = 22$	0	9 −42.1 1	9 −41.2 34	7 −50.5 2	5 −43.0 3
6	WD-97-011894	$K_i = 8.2$	0	9 −37.6 5	9 −43.8 12	9 −44.5 13	6 −40.9 6
7	WD-99-014082	$IC_{50} = 90$	0	8 −38.4 3	9 −42.7 18	9 −49.6 4	5 −40.0 8
8	WD-98-015533	$K_i = 70$	0	9 −34.6 14	9 −41.7 27	8 −48.6 5	7 −41.3 5
9	WD-2000-006814 (i)	$K_i = 3$	9 −26.9 54	9 −23.8 115	9 −34.8 101	9 −36.1 50	9 −26.0 105
10	WD-2000-006814 (ii)	$K_i = 5$	0	3* −14.2 202	9 −30.7 167	0	0
11	WD-2001-003727	$IC_{50} = 10$	0	7 −31.3 28	9 −45.5 7	4 −38.5 32	3 −33.7 37
12	WD-2001-002907	"inhibited MMP-1"	6 −35.2 7	7 −31.4 27	9 −40.0 47	7 −38.5 31	5 −35.4 31
13	WD-99-011608 (i)	$K_i = 423$	0	0	5* −36.9 74	0	0
14	WD-99-011608 (ii)	$K_i = 224$	0	0	3* −45.7 5	0	0
15	WD-99-004584	$IC_{50} = 360$	0	0	5* −36.1 85	0	0
16	WD-99-016138	$IC_{50} = 1.3$	0	7 −27.3 63	9 −36.3 83	0	0

[a] MMP-1 activities as reported in WDI and/or by Whittaker et al.[13] For each of the five ensembles/models, the hit-rate out of 10 attempts, reagent interaction score, and ranking among the successful reagents are listed. For hit rates greater than 1, the interaction score and the ranking shown are the best values identified in that set of hits. Hit rates are labeled with an asterisk (*) when the reagents were successfully screened in only one of the conformers of a particular ensemble, and zero hits were identified in all other conformers of that ensemble.

side chain conformational diversity. As discussed above, this is probably a direct consequence of the backbone variation exhibited by these models that is not present in the computational methods.

Receptor−Ligand Interaction Scores. In this work, a scoring function very similar to Böhm's empirical scoring function[36] was used as a measure of receptor−ligand interactions.[37] The scoring function is knowledge-based and includes H bonds, ionic interactions, lipophilic protein−ligand contacts, and the number of rotatable bonds in the ligand. An interaction score was computed for each successful reagent in the five model sets, and these reagents were ranked in order of the score. The rankings of virtual hits from the Dynasite ensemble and Reflex approach were plotted against that of the NMR_{Holo} ensemble as depicted in Figure 4. A medium quality correlation ($r^2 = 0.50-0.52$) was observed between the interaction score ranks of virtual hits in these models. However, the correlation between NMR_{Apo} and NMR_{Holo} was also 0.52 (data not shown), which indicates that this is the level of correlation one can expect when comparing two experimental ensembles. Reagents that were unsuccessful in the X-ray model, but identified by Dynasite, Reflex, and NMR_{Holo} models, are highlighted in Figure 4. These results compellingly suggest that there are several compounds that potentially interact strongly with the S1′ pocket of MMP-1, in which the receptor adopts an alternative conformation to the X-ray crystal structure of 2TCL.

Activity. The question remains: are the additional virtual hits with the flexible strategies, particularly those with high-scoring protein−ligand interactions, active toward MMP-1? There are several compounds reported in the literature with binding data that have the exact or very similar scaffolds compared to the one used in this study. Figure 5 depicts a selection of relevant MMP-1 inhibitors extracted from the Derwent World Drug Index (WDI) (Derwent Information Ltd., London, U.K.) with their activities listed in Table 2. The portion of these ligands corresponding to the reagents used in the screening is shown in bold. The ligands all contain the hydroxamate moiety and the sulfonyl group, the number of bonds between the two is constant, and the substituent groups are all attached to the same sulfonyl group. The compounds are expected to participate in the same H bonds (see Figure 1b), and their sulfonyl substituent is expected to occupy the same region of the S1′ pocket. This has been verified experimentally for CGS-27023A (compound **1**) in NMR structure 3AYK[19] and RS-104966 (identical to compound **8** with the terminal chlorine atom substituted for a hydrogen atom) in the X-ray structure 966C.[20] The high geometrical conservation of the substitution vectors of these two diverse hydroxamate compounds after trans-

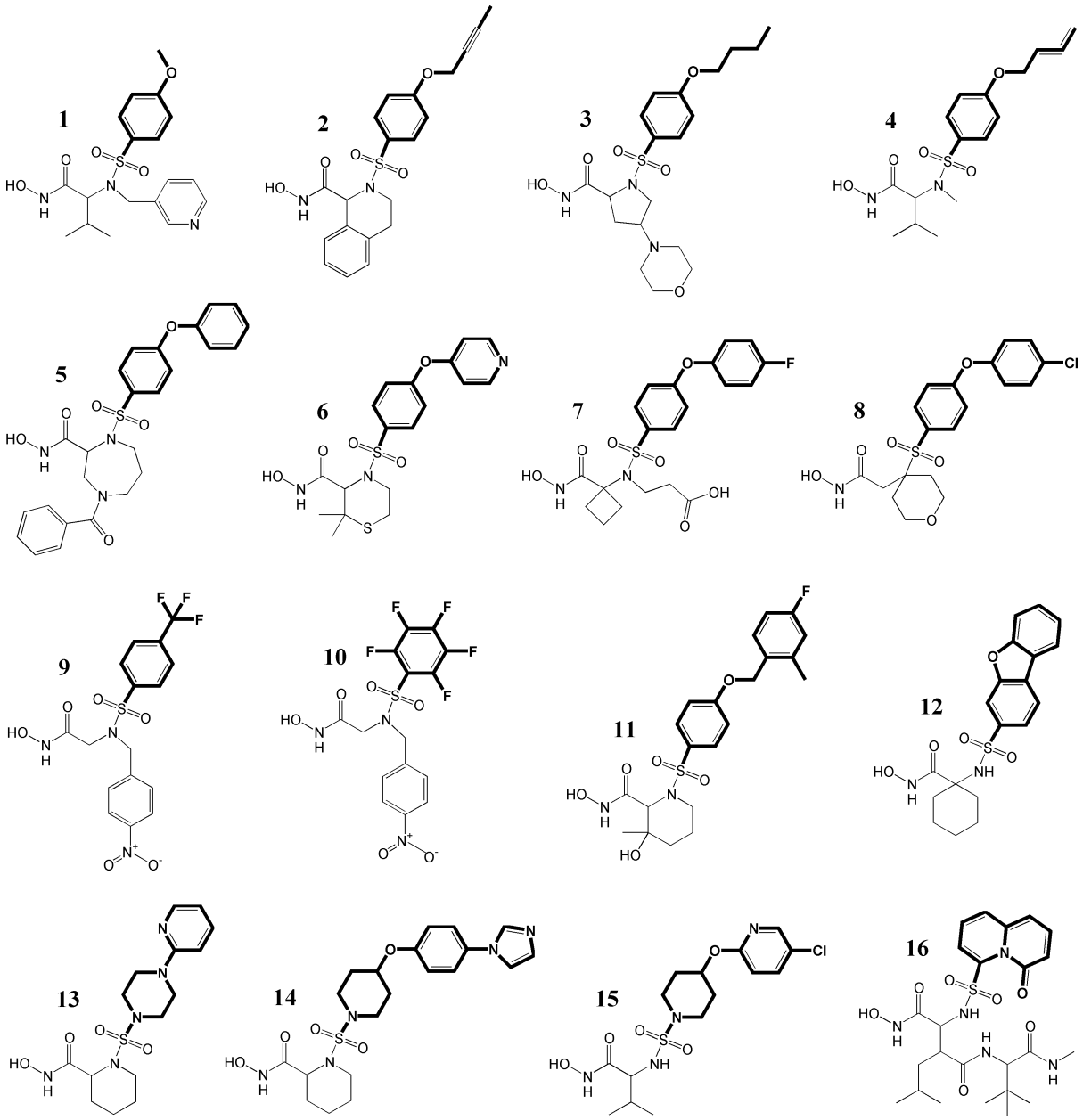

Figure 5. Selection of MMP-1 inhibitors from the WDI. See Table 2 for further information.

lation into the same coordinate framework is shown in Figure 6. For the reasons given above, we expect that this subset of reagents attached to our common scaffold will generate ligands that have binding characteristics very similar to those of the known inhibitors in Figure 5, and thus, we can relate their activities with MMP-1 to the success or failure of the reagents in the screening process.

The outcome of the screening of these 16 reagents with each of the five ensembles/models is reported in Table 2. The reagents can be subdivided into four sets broadly corresponding to increasing reagent size. The smallest reagents in the first set (**1**−**4**), comprising alkoxyphenyl groups, which yield K_i or IC_{50} values for MMP-1 ranging from 24 to 920 nM for the corresponding ligands in Figure 5 (see Table 2 for activity values), are successfully screened by all five strategies. These relatively small fragments can "fit" into the S1′ pocket even in the absence of any conformational changes. Table 2 also reports the interaction score and its rank for the corresponding reagents for all five ensembles/models. The first set of reagents does not rank particularly highly among the virtual hits for each set, although their ranking is highest among the hits for the static X-ray model.

The second set of reagents (**5**−**8**) comprise larger biaryl substituents. The corresponding ligands in Figure 5 are highly potent MMP-1 inhibitors with K_i or IC_{50} values ranging from 8.2 to 90 nM. These reagents do not "fit" into the S1′ cavity of the static X-ray model because of severe steric clashes with the site. However, they are successfully selected by the NMR models as well as by both flexible methods (Dynasite/Reflex). Rotation of the Arg214 and Phe242 side chains opens up the pocket and allows occupancy of the S1′ cavity by these reagents without steric clashes. Fragments of this type lead to highly active MMP-1 inhibitors, which would not have been identified from utilization of the static X-ray model. According to Table 2, these reagents are ranked highly in terms of the interaction score.

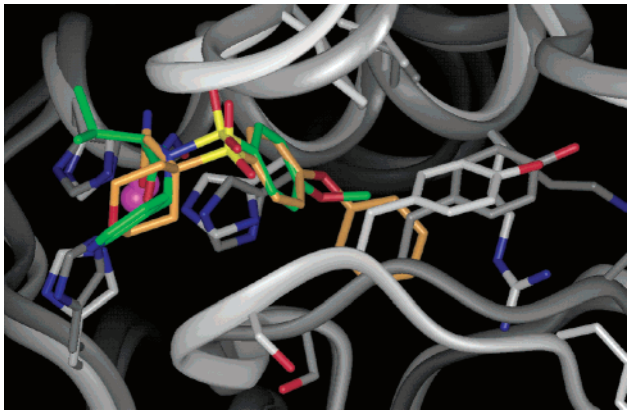

Figure 6. Backbone superposition of X-ray structures 3AYK and 966C showing the conservation of the sulfonyl substitution vector across diverse hydroxamate scaffolds. Ligands CGS-27023A (3AYK) and RS-104966 (966C) are shown with green and orange carbon atoms, respectively. The backbone and Zn-coordinating histidines, Arg214, Tyr240, and Ser239, are shown for each protein (966C in dark-gray). Zn atoms are shown in purple. 3AYK is the average structure of 4AYK.

The third set of reagents (**9**−**12**) comprises fluorinated groups, an extended biphenyl ether system, and a fused tricyclic system. The K_i or IC_{50} values of the corresponding highly potent ligands in Figure 5 range from 3 to 10 nM. Reagents **9** and **12** are successfully screened by all approaches, whereas reagent **11** with an IC_{50} value of 10 nM for the actual ligand is unsuccessful with the X-ray model but is identified as a "hit" with the NMR models and the computational methods. Reagent **10** is only successful with the NMR models. In general, the interaction score ranking for the virtual hits in this set of reagents is not very high.

The final set of reagents (**13**−**16**) comprises two- and three-ring systems, with an aliphatic ring attached to the sulfonyl substitution point in reagents **13**−**15**. The molecule in Figure 5 corresponding to reagent **16** is highly active (IC_{50} value of 1.3 nM), whereas those corresponding to reagents **13**−**15** are less active (K_i or IC_{50} between 224 and 423 nM). Reagent **16** is only screened successfully with the NMR models, and reagents **13**−**15** are only identified as virtual hits with the NMR_{Holo} target. It seems as if the binding of these reagents requires a backbone rearrangment such as that present in the NMR models (see sections above). Interestingly, only one of the NMR_{Holo} models could accommodate the latter reagents, with zero virtual hits identified for the other models in the NMR_{Holo} ensemble. The interaction score ranking was again not very high for these reagents in the NMR models.

From a consideration of the data in Table 2, it is clear that the flexible receptor approaches can successfully identify highly potent molecules that would be missed using a static receptor. This demonstrates the value of integrating binding site flexibility into the sceening process to recognize more active, diverse ligands.

Conclusions

Dynasite and Reflex have been validated as tools for conformational sampling of local side chain flexibility. The value of the algorithms for increasing the virtual hit rate and diversity coverage of reagents identified in virtual screening has also been demonstrated. These methods are alternatives to molecular dynamics for exploring the receptor conformational space prior to or during virtual screening processes. The main strength of the methods described in this work is that they are significantly faster computationally than molecular dynamics. This is particularly striking for the Reflex approach, which samples conformational space during the screening procedure.

In the screening process, several reagents that lead to highly potent MMP-1 inhibitors were successfully identified with the flexible approaches but failed with the static X-ray model. This confirms the validity of our receptor flexibility methods for recognizing active ligands in virtual screening. The NMR models also successfully recognized these as well as additional active reagents, and thus, it is feasible and computationally efficient to use NMR structures, when available, in such studies. However, there are far fewer NMR structures (3224) in the PDB than X-ray models (17 927 as of the June 3, 2003, release of the PDB), and thus, when NMR models are not available, the flexible methods presented here provide a suitable alternative. The main weakness is that they do not address backbone variations, which, in turn, could induce increased side chain diversity. Nevertheless, as the PCA analysis of the Dynasite ensemble and Reflex conformers exhibited, our flexible methods encompass a significant proportion of the side chain conformational space spanned by the NMR_{Apo} and the NMR_{Holo} models in the case of the S1′ pocket of MMP-1. Furthermore, it is known that side chain structure is less well conserved than backbone dihedral angles. Bearing this in mind, and the evidence of the results presented, our flexible binding site methods afford valid and efficient approaches for incorporating protein conformational variations into the lead identification process.

Acknowledgment. We thank Dr Philip M. Dean for many fruitful discussions concerning the methodology and Dr David T. Manallack for useful comments relating to MMP inhibitors.

References

(1) Luong, C.; Miller, A.; Barnett, J.; Chow, J.; Ramesha, C.; Browner, M. F. Flexibility of the NSAID Binding Site in the Structure of Human Cyclooxygenase-2. *Nat. Struct. Biol.* **1996**, *3*, 927−933.
(2) Cao, Y.; Musah, R.; Wilcox, S.; Goodin, D.; McRee, D. Protein Conformer Selection by Ligand Binding Observed with Crystallography. *Protein Sci.* **1998**, *7*, 72−78.
(3) Carlson, H. A. Protein Flexibility and Drug Design: How to Hit a Moving Target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(4) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular Docking of Ensembles of Structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(5) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(6) Ota, N.; Agard, D. A. Binding Mode Prediction for a Flexible Ligand in a Flexible Pocket Using Multiconformation Simulated Annealing Pseudo Crystallographic Refinement. *J. Mol. Biol.* **2001**, *314*, 607−617.
(7) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuation of a Protein in Computer-Aided Drug Design. *J. Phys. Chem.* **1999**, *103*, 10213−10219.
(8) Schnecke, V.; Kuhn, L. A. Virtual Screening with Solvation and Ligand-Induced Complementarity. *Perspect. Drug Discovery Des.* **2000**, *20*, 171−190.
(9) Anderson, A. C.; O'Niel, R. H.; Surti, T. S.; Stroud, R. M. Approaches to Solving the Rigid Receptor Problem by Identifying a Minimal Set of Flexible Residues during Ligand Docking. *Chem. Biol.* **2001**, *8*, 445−457.
(10) Leach, A. R. Ligand Docking to Proteins with Discrete Side-Chain Flexibility. *J. Mol. Biol.* **1994**, *235*, 345−356.

(11) Caflisch, A.; Miranker, A.; Karplus, M. Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV-1 Aspartic Proteinase. *J. Med. Chem.* **1993**, *36*, 2142−2167.
(12) Teague, S. J. Implications of Protein Flexibility for Drug Design. *Nat. Rev. Drug Discovery*. **2003**, *2*, 527−541.
(13) Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. H. Design and Therapeutic Application of Matrix Metalloproteinase Inhibitors. *Chem. Rev*. **1999**, *99*, 2735−2776.
(14) Coussence, L. M.; Fingleton, B.; Matrisian, L. M. Matrix Metalloproteinase Inhibitors and Cancer: Trials and Tribulations. *Science* **2002**, *295*, 2387−2392.
(15) Cleutjens, J. P. The Role of Matrix Metalloproteinases in Heart Disease. *Cardiovasc. Res*. **1996**, *30*, 816−821.
(16) Beckett, R. P.; Whittaker, M. Matrix Metalloproteinase Inhibitors. *Expert Opin. Ther. Pat.* **1998**, *8*, 259−282.
(17) Vassiliou, S.; Mucha, A.; Cuniasse, P.; Georgiadis, D.; Lucetlevannier, K.; Beau, F.; Kannan, R.; Murphy, G.; Knauper, V.; Rio, M. C.; et al. Phosphinic Pseudo-Tripeptides as Potent Inhibitors of Matrix Metalloproteinases: a Structure−Activity Study. *J. Med. Chem*. **1999**, *42*, 2610−2620.
(18) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97*, 1359−1472.
(19) Moy, F. J.; Chanda, P. K.; Chen, J. M.; Cosmi, S.; Edris, W.; Skotnicki, J. W.; Powers, R. Solution Structure of the Catalytic Fragment of Human Fibroblast Collagenase Complexed with a Sulfonamide Derivative of a Hydroxamic Acid Compound. *Biochemistry* **1999**, *38*, 7085−7096.
(20) Lovejoy, B.; Welsh, A. R.; Carr, S.; Luong, C.; Broka, C.; Hendricks, R. T.; Campbell, J. A.; Walker, K. A. M.; Martin, R.; van Wart, H.; et al. Crystal Structures of MMP-1 and -13 Reveal the Structural Basis for Selectivity of Collagenase Inhibitors. *Nat. Struct. Biol.* **1999**, *6*, 217−221.
(21) Botos, I.; Scapozza, L.; Zhang, D.; Liotta, L. A.; Meyer, E. F. Batimastat, a potent matrix mealloproteinase inhibitor, exhibits an unexpected mode of binding. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2749−2754.
(22) Meng, Q.; Malinovskii, V.; Huang, W.; Hu, Y.; Chung, L.; Nagase, H.; Bode, W.; Maskos, K.; Brew, K. Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1′ residue of substrate. *J. Biol. Chem*. **1999**, *274*, 10184−10189.
(23) Borkakoti, N.; Winkler, F. K.; Williams, D. H.; Darcy, A.; Broadhurst, M. J.; Brown, P. A.; Johnson, W. H.; Murray, E. J. Structure of the Catalytic Domain of Human Fibroblast Collagenase Complexed with an Inhibitor. *Nat. Struct. Biol.* **1994**, *1*, 106−110.
(24) Moy, F. J.; Chanda, P. K.; Cosmi, S.; Pisano, M. R.; Urbano, C.; Wilhelm, J.; Powers, R. High-Resolution Solution Structure of the Inhibitor-Free Catalytic Fragment of Human Fibroblast Determined by Multidimensional NMR. *Biochemistry* **1998**, *37*, 1495−1504.
(25) Källblad, P.; Dean, P. M. Efficient Conformational Sampling of Local Side-Chain Flexibility. *J. Mol. Biol*. **2003**, *326*, 1651−1665.
(26) Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. **2000**, *28*, 45−48.
(27) Maple, J. R.; Dinur, U.; Hagler, E. T. Derivation of Force Fields for Molecular Mechanics and Dynamics from ab Initio Energy Surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5350−5354.
(28) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins−Application to Side-Chain Prediction. *J. Mol. Biol*. **1993**, *230*, 543−574.
(29) Dunbrack, R. L.; Cohen, F. E. Bayesian Statistical Analysis of Protein Side-Chain Rotamer Conformations. *Protein Sci*. **1997**, *6*, 1661−1681.
(30) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem*. **1968**, *68*, 441−451.
(31) Jackson, J. E. *A User's Guide to Principal Components*; Wiley: New York, 1991.
(32) MacPerson, L. J.; Bayburt, E. K.; Capparelli, M. P.; Carroll, B. J.; Goldstein, R.; Justice, M. R.; Zhu, L.; Hu, S.; Melton, R. A.; Fryer, L.; et al. Discovery of CGS 27023A, a Non-Peptidic, Potent, and Orally Active Stromelysin Inhibitor That Blocks Cartilage Degradation in Rabbits. *J. Med. Chem*. **1997**, *40*, 2525−2532.
(33) Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Böhm, H.-J.; Dean, P. M. A Validation Study on the Practical Use of Automated de Novo Design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 459−478.
(34) Todorov, N. P.; Dean, P. M. Evaluation of a Method for Controlling Molecular Scaffold Diversity in de Novo Ligand Design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175−192.
(35) Schrauber, H.; Eisenhaber, F.; Argos, P. Rotamers: to be or not to be? An Analysis of Amino-Acid Side-Chain Conformations in Globular Proteins. *J. Mol. Biol*. **1993**, *230*, 592−612.
(36) Böhm, H.-J. The Development of a Simple Empirical Scoring Function To Estimate the Binding Constant for a Protein−Ligand Complex of Known Three-Dimensional Structure. *J. Comput.-Aided Mol. Des*. **1994**, *8*, 243−256.
(37) Todorov, N. P.; Dean, P. M. A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 335−349.

JM031061L