Tissue components hydrolyzing matrix metalloproteinases (MMPs) exhibit a high sequence similarity (56–64% in catalytic domains) and yet a significant degree of functional specificity. The hexapeptide-binding sites of 24 known human MMPs were compared in terms of their force field interaction energies with five probes that are most frequently encountered in substrates and inhibitors. The probes moved along a grid enclosing partially flexible binding sites in rigid catalytic domains that were represented by published experimental structures and comparative models and new comparative models for nine most recently characterized MMPs. For individual MMPs, representative interaction energies were obtained as averages for all suitable experimental structures. Correlations of the representative energies for all MMP pairs were succinctly catalogued for individual probes, subsites, and correlation levels. Among the probes (neutral sp$^3$ carbon and sp$^2$ oxygen, positive sp$^2$ nitrogen and hydrogen, and negative carbonyl oxygen), the last probe is least distinctive. Similarities of subsites are decreasing as S1$'$ > S2 > S3$'$ > S1 – S3 > S2. Most interesting, occurrences of subsites in published complexes of MMP-inhibitor complexes follow an almost parallel trend, alluding to overall low selectivity of known MMP inhibitors. Flexible subsite S1$'$ that appears as the specificity pocket in rigid x-ray structures is actually very similar among individual MMPs. Several correlations indicated that MMPs 3, 8, and 12 have similar binding sites. Modeling results are corroborated with published experimental data on MMP inhibition and substrate specificities. The results provide numerous clues for development of specific inhibitors and substrates, as well as for selection of MMPs for testing that provides maximum information without redundant experiments.

Matrix metalloproteinases (MMPs$^*$) facilitate remodeling of connective tissues by hydrolysis of extracellular matrix components. The MMP degradation potential is controlled at the expression and activation levels, as well as by specific inhibitors (TIMPs)$^{(1, 2)}$. In various pathological conditions including inflammation, arthritis, and cancer, specific MMPs manifest imbalanced activities, which are targeted by drug development projects. This study intends to contribute to such projects by analysis of structural similarities of binding sites of MMPs, with the aim to provide a basis for learning which MMPs can be selectively inhibited, which features confer specificity on inhibitors, and which MMPs are to be tested to avoid redundancy.

Catalytic domains of MMPs possess high sequence similarity (56–64%). A conserved sequence HELGHXXGXXHI in the C-terminal portion is coordinating catalytic zinc atom via the three His residues (3). MMPs 2 and 9 have catalytic domain split into two parts by insertion of a fibronectin-like domain, but this insert does not seem to change the overall fold (4). Traditionally, S1$'$ subsite (5) has been labeled as the specificity pocket (6, 7) because of its significant size and shape differences among the rigid x-ray structures of various MMPs. However, modifications of P1$'$ part of the ligands led to both increased selectivity (8, 9) and failures (10, 11) that were explained by observed changes in the structure of the pocket upon inhibitor binding (12–18). The importance of other parts of the binding site, especially the unprimed side (11, 19–21), for the design of more potent and selective inhibitors has been demonstrated as well. Occupation of multiple subsites may be crucial for designing truly specific inhibitors (21).

The possibility of induced fit makes determination of similarity of catalytic domains based solely on their specificity to macromolecular substrates, sequence similarity, or the similarity of rigid structures elusive. More insight is expected from examination of structural changes induced by ligand binding and the energies of the resulting complexes. The methods using the force field interaction energies of probes in individual points of a grid positioned in the binding site have a potential to provide more general results than binding studies of concrete inhibitor structures (22), especially if the binding site is treated as flexible to account for induced fit (23). The approach was used to search for the probes that have different interaction energies in individual subsites of MMPs (23). The results provide valuable guidelines for design of selective inhibitors.

This study aims at a more detailed classification of MMPs in terms of similarity of binding sites. For this purpose, the interaction energies of probes in flexible binding site are perfectly suitable. The previously used (22, 23) principal component analysis (PCA) or consensus PCA (CPCA) requires omission of the repulsive interaction energies. Because the repulsive energies are extremely important in size and shape definitions of the binding site, we compared all interaction energies of all MMP pairs directly, by linear regression analysis (LRA). Although more time-consuming than CPCA, the method precisely quantifies all similarities instead of focusing on the most pronounced differences. The study is based on a complete set of 24

---

* The abbreviations used are: MMPs, matrix metalloproteinases; CPCA, consensus principal component analysis; PCA, principal component analysis; LRA, linear regression analysis; PDB, Protein Data Bank; S3, S2, S1, S1$'$, S2, S3, MMP subsite binding amino acid residues of a hexapeptide stretch of a substrate on both sites of the S1-binding amino acid that has the peptide-bond carbonyl coordinated with catalytic zinc; TIMP, tissue inhibitor of MMP.
currently known human MMPs (x-ray or NMR structures for 9 MMPs and 15 comparative models) as compared with 10 MMPs (x-ray structures for MMPs 1–3, 7, and 8; and comparative models for MMPs 9, 12–14, and 20) used previously (23).

MATERIALS AND METHODS

Protein Structures—X-ray and NMR structures of human MMPs, with or without ligands, were obtained from the Protein Data Bank (PDB) files (24) as follows: MMP1: 1AYK, 1CGL, 1HFC, 2AYK, 2TCL, 3AYK, 4AYK, 966C, 1CGE, 1CGF; MMP2: 1HOV, 1QBP; MMP3: 1B3D, 1BSY, 1BIW, 1BQO, 1C3I, 1CAQ, 1CIZ, 1CQR, 1D5J, 1D7X, 1DF5, 1DOH, 1G4F, 1G4K, 1GOS, 1HFS, 1HYY, 1LSM, 1LSN, 1UEA, 1UMT, 1UMT, 1UNQ, 2USN, 1QIA, 1QIC; MMP7: 1MPM, 1MKQ, 1MMR; MMP8: 1A85, 1A86, 1B2S, 1J73, 1J76, 1JAN, 1JAO, 1JAP, 1JQQ, 1JHI, 1J9J, 1KBC, 1MBY, 1MNC; MMP9: 1GKC; MMP12: 1LZJ, 1JK3; MMP13: 1EUB, 1FLS, 1FM1, 456C, 830C; MMP14: 1BQQ, 1BUV. In the NMR files (shown in italics) containing multiple structures, only the first structure was selected. If the files contained multiple domains or multimers of the same domain, only one domain with the catalytic site was used. All experimental structures were modified by removing the water molecules and ligands and by adding hydrogens (25). Comparative models of MMPs 2 and 9–18 were obtained from Mobashery and co-workers (26).

Comparative Models—Comparative models for the remaining nine human MMPs, and 26–28 were generated (27) using the PDB files 96542, 606882, QSN119, QUBR9, QYSR2, QNPA2, Q9H536, and Q9H329 from the Swiss-Prot data base (28). The x-ray structures served as reference structures; if an MMP has several x-ray structures available in the PDB, the structure with the best resolution was selected. The structures of catalytic domains were aligned by using the conserved regions in individual subunits as well as all other regions with 100% identity throughout all structures (29, 30). For each modeled structure, a primary reference structure was selected based on the highest identity score (modeled MMP-reference MMP) as follows: 19–1, 20–2, 21–1, 23–2, 24–1, 25–9, 26–8, 27–2, and 28–1. The sequences of modeled proteins were aligned one by one to the sequences of primary reference proteins based on the conserved regions. The coordinates for residues within conserved regions were assigned to each modeled structure from the corresponding primary reference. Then the coordinates for the sequences at least three residues long, identical to at least one of the reference structures, were directly assigned. If a sequence was identical to more than one reference structure, the priority was set as follows: (i) the primary reference structure, (ii) the reference structure where more identical regions were identified, and (iii) other reference structures. For non-conserved residues, the coordinates were generated de novo (27) by using the following parameters: convergence 0.05, closure iterations 1000, internal overlap 0.5, external overlap 0.8, and scale torsions 60. From generated loops, the one that had the backbone closest to the backbone of the reference protein was selected. The coordinates for end residues were assigned using a standard procedure (27).

Steric overlaps were relieved by (i) re-assigning the coordinates for conflicting residues from the same reference for conserved residues and (ii) changing one of the torsions in the side chain for the residues with the de novo coordinates. The spacers were repaired by optimization (31). For de novo positioned residues, the initial extended conformations of the side chains were modified by using a rotamer search (31). The zinc and calcium atoms were added using the coordinates in the primary reference structure.

Final optimization using the vff force field (31) with the native charges on the atoms and the charge +2 on zinc atoms proceeded iteratively in several phases. First, only the side chains of residues within the conserved regions and directly assigned loops that were different from reference structures were optimized. In the second phase, optimization of the de novo generated loops was performed, including both the backbone and the side chains. The regimes for both phases were identical. The first cycle consisted of 100 minimization iterations (steepest descent gradient, derivative 0.001) and 1000 dynamics steps (equilibration 100 steps, temperature 300 K). For the following 10 cycles, the steepest descent gradient was changed to conjugate gradient and cross-terms were added to both minimization and dynamics. The dynamics trajectory was not showing significant energy fluctuations within final conformation of protein with dynamics refinement starting conformation for the next cycle. The third phase optimized conformations of six His residues coordinating zinc atoms, with template forcing of the ring atoms. Minimization alone was used as follows: 100 iterations with steepest descent gradient were followed by 500 iterations with conjugate gradient. In the final phase (100 minimization iterations with conjugate gradient), the entire structure was optimized, with template forcing of imidazole rings of zinc-coordinating His residues. All four optimization phases were repeated once, with a slight change in phases one and two (only 100 dynamics iterations were performed in each cycle and the number of cycles with conjugate gradient was set to five).

Validation of Comparative Modeling Procedure—The comparative model was created for one of the MMPs (MMP 2) with available x-ray structure and a comparative model prepared by a different group (26) to assess consistency between the three structures. The root mean square deviation between the x-ray structure and our comparative model was 1.614 Å compared with 1.564 Å for the published comparative model (26). The overall backbone alignment among the three structures was tight, with the exception of two loops, which reside on the surface of the protein, in a sufficient distance from the binding site to preclude any interactions with the binding site residues or the probe. In both comparative models, the residues forming individual subunits of the binding site exhibited conformations that were very close to those observed in the x-ray structure. Probe interaction energies (the details are given below) for all three structure pairs were closely correlated. Good agreement of our and published (26) comparative models, both mutually and with the x-ray structure, warrants their use for the MMPs where no experimental structures (x-ray or NMR) were available.

Calculation of Interaction Energies—All used MMP catalytic domain structures (23) were aligned (27) using the PDB file 1A85 (one of the x-ray structures for MMP-8) as a template. Binding sites consisting of 36 residues and catalytic zinc were defined based on the literature (29, 34). An irregular grid consisting of 1770 points with spacing 1.5 Å was constructed to completely enclose the binding site (Fig. 1). This arrangement was preferred to a regular box in order to minimize the number of points where to calculate the interaction energies. The grid points that were placed beyond the residues lining the binding site, inside the protein, were excluded. The remaining 941 grid points were associated with the subsites (5) based on the published data (29, 34), with subsites S1, S2...S3 represented by 245, 179, 147, 129, 126, and 115 points, respectively. Five probes representing the atom types with highest probability of occurrence in small substrates and inhibitors were selected: neutral sp3 oxygen and sp3 carbon atoms, sp2 nitrogen atom with charge +1, hydrogen atom with charge +1, and carbonyl oxygen atom with charge −1.5.

The probe was placed in each grid point (Fig. 1); geometry of the complex was optimized, and the total energy was calculated using the electrostatic and van der Waals terms of the Tripos force field (25), with the Gasteiger charges (33) on protein atoms and the charge +2 on zinc atoms. For geometry optimization, the amino acids that had at least one atom within 2 Å distance from the probe were set as flexible; the rest of the protein and the position of the probe were fixed. The steepest descent method with the termination criterion of the energy change less than 0.5 kcal/mol and maximum of 1000 iterations was used for geometry optimization. Interaction energy was calculated as the difference between energy of optimized protein-probe complex and the single-point energy of optimized protein without the probe.

Representative Sets of Interaction Energies—All available experimental structures for each MMP were used to calculate the interaction energies to ensure robust comparisons that are based on the differences in MMPs themselves, rather than on the differences caused by various ways of obtaining individual structures. For each MMP, a representative set was created as an average of the interaction energies in each grid point for similar structures. Two structures were considered similar, if the linear correlation of their interaction energies explained at least 95% of variance (R2 ≥ 0.7, where R is the correlation coefficient). For MMPs with only two available structures or with more than two structures but no similar structures, all structures were included in the average. For MMPs with multiple groups of similar structures, the group including more structures originating from different laboratories was chosen to create the representative set of interaction energies.

RESULTS AND DISCUSSION

The properties of binding sites of various MMPs were compared using the interaction energies with five probes representing the atom types that are most frequently occurring in
Similarity of Human MMP-binding Sites

Fig. 1. Grid enclosing the binding site and its parts belonging to individual subsites. Catalytic zinc shown as a sphere. Catalytic domain of MMP 8 (PDB entry 1A85) is shown; tubes indicate amino acids (149DNSPFD154, 157NGILAHAF164, 191LFLVAAHE198, 201HSL-)

the peptidic substrates and non-peptidic inhibitors of MMPs (neutral sp² carbon and sp³ oxygen, positive sp³ nitrogen and hydrogen, and negative carbonyl oxygen). The used protein structures included experimental structures from the PDB, published comparative models (26), and comparative models for nine most recently characterized MMPs, which were prepared in this study.

Probe Interaction Energies in the MMP-binding Sites—The binding site was enclosed in an irregular lattice of grid points taking actual contours of the binding site (Fig. 1). The probe was placed in each of the grid points, and the interaction energy was calculated with the partially flexible binding site in order to account for experimentally observed induced fit (12). The approach also optimizes the backbone parts of amino acids, in contrast to a previous study (23) where only the side chains were movable. An analysis of the temperature factors for MMP structures in the PDB shows that there are no substantial differences in mobility of the side chains and the backbone in the MMP-binding sites (data not shown). For individual MMPs, the representative sets of interaction energies, obtained by averaging the interaction energies for all suitable experimental structures, were used in all comparisons. The pairwise comparisons were performed by LRA. Because the fitted lines did not substantially differ from the identity line, explained variance (R²) was used to express the level of similarity.

MMP Similarity—A comparison of the representative interaction energies for all probes in the entire binding sites of all 276 (24 × 23/2) pairs of known human MMPs (Fig. 2) revealed several groups of MMPs, which are similar at fairly high levels. The interaction energies showed the highest correlation between entire binding sites for MMPs 3 and 7 (R² = 0.9). At the level R² ≥ 0.8, five groups of MMPs were identified: (i) MMPs 2, 3, 8, and 12; (ii) MMPs 3, 7, 8, and 12; (iii) MMPs 1, 8, and 13; (iv) MMPs 3, 7, 12, and 14; and (v) MMPs 3, 8, 12, and 13. Seven groups of MMPs were shown to be similar at the level R² ≥ 0.7: (i) MMPs 1–3, 7, 8, and 12–14; (ii) MMPs 2, 3, 7–9, 12, and 14; (iii) MMPs 3, 7, 10, 12, and 14; (iv) MMPs 12, 14, and 16; (v) MMPs 12, 14, and 18; (vi) MMPs 12, 17; and (vii) MMPs 11 and 18.

Fig. 2. Pairwise correlations of interaction energies for complete binding sites of all 24 MMPs and probes. The upper triangular matrix contains R² ranges for all probes, the lower matrix for neutral probes. The colors indicate the correlation level (R²): 0.90–1.00 (red box), 0.80–0.89 (amber box), 0.70–0.79 (tan box), 0.60–0.69 (gold box), 0.50–0.59 (yellow box), 0.40–0.49 (white box), 0.30–0.39 (aqua box), 0.20–0.29 (light blue box), 0.10–0.19 (medium blue box), 0.00–0.09 (dark blue box).

Generally, the charged probes exhibit wider ranges of interaction energies and, hence, a larger impact on overall correlations than the neutral probes. The situation for uncharged probes is also depicted in Fig. 2. In a summary for neutral probes, one group of MMPs was found to be similar at the level R² ≥ 0.8 (MMPs 8 and 12) and two groups showed similarity at the level R² ≥ 0.7 (MMPs 1 and 8; and MMPs 3, 8, and 12) in the entire binding site. MMPs 1–18 exhibit higher similarity in whole binding sites than MMPs 19–28, both mutually and to MMPs in the latter group, especially if all probes are considered. MMPs 19–28 were represented by newly constructed comparative models. Because the construction procedure was validated, we believe that the lower similarity in the latter group is due to more pronounced differences (e.g., average sequence identity of MMPs 19–28 is 43.5% as compared with 51.0% among MMPs 1–18).

Subsite Similarity—To compare the subsites, interaction energies for all five probes were correlated for all 276 MMP pairs simultaneously (the energies for MMP with lower numbers were plotted on the x axis and those for MMP with higher numbers on the y axis). The overall trend, in order of decreasing similarity of subsites, is as follows: S1′ > S2 > S3 > S1, as indicated by the R² values for all probes, 0.31, 0.29, 0.26, 0.18, 0.18, and 0.16, respectively. Subsite S1′ exhibits the highest levels of similarity, followed by S2 and S3′ subsites. The results add to multiple experimental proofs of flexibility of S1′ subsite (12–18), help explain cases where desired specificity was not achieved upon optimization of groups fitting into S1′ subsite (10, 11), and contradict the view of S1′ subsite as the specificity pocket of MMPs that is based on variability of rigid x-ray structures.

An interesting picture is emerging from a comparison of subsite similarity with actual binding modes of 55 inhibitors for which the PDB files of complexes with MMPs are available. In this group of inhibitors, occupancies of individual subsites are as follows: S1′, 49; S3′, 29; S1, 20; S2′, 7; S2, 2; and S3, 1.
The most frequently occupied subsites S1’ and S3’ are among the most similar subsites; the least similar subsites S2’ and S3 are seldom engaged in binding. Among the 55 inhibitors, 21 bind in subsites S1/S3, 8 in S1/S1’, 5 in S1’, and groups of 1–3 inhibitors in other subsite combinations. Generally speaking, if the MMP inhibitors in the PDB are indicative of overall synthetic trends, the development of MMP inhibitors apparently does not target the subsites that convey specificity. It should be noted, however, that the overall subsite similarity trends are too fuzzy to be used for the design of concrete inhibitors and substrates. For this purpose, the subsites where the binding is anticipated and probes that occur in the studied compounds should be used.

**Probe Selectivity**—For comparison of individual probes, their interaction energies for all 276 MMP pairs were correlated simultaneously. The negative carbonyl oxygen probe (correlations for individual subsites in Fig. 4, 1st row) showed the least discrimination, with \( R^2 = 0.44 \) across all MMP-binding sites. This phenomenon is most pronounced in subsites S2, S3, and S1’ \( (R^2 = 0.52, 0.49, \) and 0.45, respectively). Because subsites S1’ and S2 are directly linked to catalytic zinc, could the interactions with the positively charged zinc atom overshadow the interactions with the amino acids in individual subsites? The chance is slim because: (i) not all MMPs were similar in these two subsites; (ii) S3’ subsite exhibits similar \( R^2 \), and (iii) S1 and S3 subsites, which are located close to the positively charged structural zinc atom, are showing much larger selectivity \( (R^2 = 0.30 \) and 0.26, respectively) than S1’ and S2 subsites.

Neutral probes \( (sp^3 \) carbon and \( sp^3 \) oxygen atoms; Fig. 4, 2nd and 3rd rows) and positively charged probes \( (sp^3 \) nitrogen and hydrogen atoms; Fig. 4, 4th and 5th rows) display much larger selectivity than the negative carbonyl oxygen probe. In fact, their overall discrimination is almost identical, as indicated by the \( R^2 \) values for whole binding sites 0.15, 0.15, 0.18, and 0.19, respectively.

**Similarities of Individual Subsites for Individual Probes**—The overall correlations across whole binding sites (Fig. 2) help with understanding the general similarity of MMPs. They are useful, too, for assessing specificity of small substrates, which usually bind in all six subsites. However, to estimate selectivity of a series of inhibitors that only occupy a few subsites, the comparisons based on the occupied subsites and probes that are parts of inhibitor structures are more suitable. Correlations for individual subsites widely varied, although the slopes and intercepts of all correlation lines did not substantially differ from the identity line (Fig. 3). The results for all MMPs, each probe, and each subsite are summarized in Fig. 4 that shows the ranges for all 8,280 \( R^2 \) values (listed in the Supplemental Materials).

The correlation data summarized in Fig. 4 can be used to estimate a correlation of binding affinities of inhibitors to a pair of MMPs. For this purpose, a joint \( R^2 \) value needs to be calculated for involved subsites and probes. The task corresponds to the calculation of the joint \( R^2 \) value for a linear dependence composed of several data sets with known \( R^2 \) values that differ in the scatter of data but not in the slopes and intercepts of the correlation lines. In addition to the \( R^2 \) values, the average interaction energy values and their standard deviations for each subsite and probe are necessary. The data and a description of the procedure are given in the Supplemental Material.

A view of the relatedness of individual MMPs can be obtained by their grouping according to the similarities of individual subsites (Fig. 4). The results for all probes and all subsites are summarized in Table I; similar tables can be constructed for individual probes or their groups. Rows refer to similarity levels characterized by the minimum \( R^2 \) values for the correlation between interaction energies of all pairs of listed MMPs. In most cases, several groups of similar MMPs at the given level in a particular subsite were identified. Frequently, these groups only differ in a few MMPs. For brevity, the MMPs that are similar in multiple groups are typed in boldface and alternating (groups of) MMPs are listed in brackets and/or separated by slash. For instance, for subsite S2’ at the similarity level \( R^2 \geq 0.7 \), four groups of similar MMPs have emerged as follows: (i) MMPs 3, 7, and 8; (ii) MMPs 7–9; (iii) MMPs 3, 8, and 12; and (iv) MMPs 8, 9, and 12.

Each probe identified several groups of MMPs that were similar in each subsite on the lowest similarity level in Table I \( (R^2 \geq 0.6) \). One group of MMPs has been found as similar in all subsites by all probes: MMPs 3, 8, and 12. Another group of MMPs \( (3, 7, \) and 12) is almost similar in the entire binding site based on the fact that only the neutral \( sp^3 \) carbon probe did not have correlation between interaction energies of MMPs 7 and 12 better than \( R^2 \geq 0.6 \) in a single subsite (S1’).

The results are in a partial agreement with published study (23) that classified as similar: (i) S3’ subsite in MMPs 3, 7, and 12; (ii) S2’ subsite in MMPs 3 and 7; (iii) S1’ subsite in MMPs 8 and 12, and in MMPs 1 and 7; and (iv) S3 subsite in MMPs 1 and 8. Table I and Fig. 4 show that we also have identified these groups in respective subsites at higher levels of similarity. A more detailed comparison of the results is precluded by different aims of these two studies. The former study (23) focused on selectivity issues by listing the features that distinguish a certain MMP from the rest, without discussing the similarity extent of the rest of the MMPs. Some variation in the results obtained by the two approaches can be explained by the differences in used methods. For instance, the previous study lists no significant differences among the 10 used MMPs (MMPs 1–3, 7–9, 12–14, and 20) in S1 subsite and very small differences in the S2’ subsite. According to our data (Fig. 4), these two subsites, along with S3 subsite, seem to be relatively different as indicated by the average \( R^2 \) values: 0.55 ± 0.24, 0.54 ± 0.28, and 0.55 ± 0.24 for subsites S1, S2’, and S3 compared with 0.67 ± 0.20, 0.70 ± 0.23, and 0.69 ± 0.28 for subsites S1’, S2, and S3’, respectively. The analysis of subsite similarity using all probes across all MMPs (see below) also showed that S1, S2’, and S3 subsites are less similar than the rest. This discrepancy may be caused by the difference in the alignments that were based on the backbones of the entire

![Fig. 3. A plot of representative interaction energies for MMPs 1 and 12 with neutral sp³ carbon probe, for the whole binding site and individual subsites.](http://www.jbc.org)
The IC_{50} values for the set of 90 arylsulfonyl-1,2,3,4-tetrahydro-isoquinoline-3-carboxylates and -hydroxamates (22, 35) were measured against MMP 3 and MMP 8. The correlation between the IC_{50} values for these two enzymes is characterized by R^2 = 0.83. The authors suggested that the inhibitors mainly bind in S1’ subsite. The overall correlation of interaction energies for all five probes in S1’ subsite for MMPs 3 and 8 is R^2 = 0.84.

A homogeneous set of 67 sulfonlated hydroxamates and carboxylates (9) features the N-p-nitrobenzyl-glycine skeleton with various functional groups filling S1’ subsite of MMPs. The p-nitrobenzyl group interacts with the S2’ subsite. The reported K_i values for MMPs 2, 8, and 9 are closely correlated (R^2 = 0.93 for MMP 2 versus both MMPs 8 and 9; R^2 = 0.91 for MMP 8 versus MMP 9). In these correlations, all 64 compounds that have one zinc-binding group were considered, to avoid ambiguity in binding modes. The structures of the compounds suggest that they should bind in subsites S1’ and S2’, using neutral and positive atoms in S1’ subsite and neutral atoms in S2’ subsite. The overall correlations between the interaction energies for these combinations of subsites and probes are in the range R^2 = 0.78–0.83 for MMPs 2, 8, and 9. Identical correlations between interaction energies for MMP1 and MMPs 2, 8, and 9 are in the range R^2 = 0.63–0.68. The comparison of the K_i values between MMP1 and MMPs 2, 8, and 9 showed little or no correlation.

The interaction energy analysis shows that MMPs 2 and 3 are highly similar in the subsites S1’ and S3’ (Fig. 4, R^2 ≥ 0.8 for each probe and subsite). The results are corroborated by correlation (R^2 = 0.72) of K_i values measured for the set of 64 carboxalkyl peptides and thiadiazole urea compounds (37, 38). A slight scatter in the data and deviation from identity line might be caused by the fact that some of the larger compounds could interact not only with the similar subsites S1’ and S3’ but also with the subsites S1 and S2’, which are by several probes identified as similar only at the level R^2 ≥ 0.6. However, the overall correlation obtained using interaction energies for probes relevant to binding of the given sets of compounds within all four subsites is R^2 > 0.8.

The interaction energies indicate that MMPs 3 and 9 are fairly similar at the primed side of the binding site (Fig. 4). This observation is supported by experimental K_i values for the set of 35 phosphonamide hydroxamates (39). The structures of the compounds suggest that neutral and positive atoms bind in S1’ subsite and neutral atoms dock into S3’ subsite. The overall correlations between interaction energies of MMPs 3 and 9 considering the relevant probes in S1’ and S3’ subsites are in...
the range $R^2 = 0.8-0.9$. The correlation between the $K_i$ values for these two enzymes is $R^2 = 0.79$. The correlations of $K_i$ for MMP 1 versus MMP 3 and MMP 9 are significantly lower ($R^2 = 0.11$ and 0.26, respectively). The correlations between calculated interaction energies of MMP 1 and MMPs 3 and 9 considering the relevant probes and subsites (Fig. 4) are characterized by $R^2$ in range 0.6–0.7.

Binding affinity includes the enthalpy component that is directly reflected in the interaction energies and the entropy component that may be, for a limited series of compounds, either invariant or decreasing with increasing interaction energies. If the similarity of interactions energies for two MMPs falls below certain limit, the variation that affects both enthalpy and entropy components may lead to a rapid decrease in correlations of experimental potencies. All the aforementioned examples adhere to a pattern that correlations of pertinent interaction energies $R^2 = 0.8$ and better are required for a good correlation of inhibitory potencies against two MMPs. The potency correlations break down completely when the correlation between interaction energies decrease into the range $R^2 = 0.6–0.7$ and below. The data also underline the importance of using the relevant probes and relevant subsites. For instance, the whole-site correlations (Fig. 2) indicate similarity of MMP 1 with MMPs 2, 3, and 8 ($R^2 \geq 0.7$), but two sets of experimental activities (9, 39) show no significant correlation between MMP1 and the remaining three enzymes.

Numerous examples exist of little or no correlation between the experimental activities of the set of inhibitors against various MMPs in accordance with our observations that the enzymes in question do not show significant similarity in the subsites occupied by the tested set of inhibitors (20, 40).

### MMP Similarities Versus Substrate Specificities—Substrates usually interact with all subsites of the binding site. Our finding that MMPs 1–3, 7, and 8 and MMPs 2, 3, and 7–9 are significantly similar in entire binding site ($R^2 \approx 0.7$, Fig. 2) is supported by the results of an extensive study on substrate specificities (41–43). The replacement of certain amino acids within the model peptidic substrate affects these enzymes in a similar way. For most pairs of MMPs from these two groups, more than 70% of replacements had the same effect on the rate of hydrolysis. The only exceptions were pairs of MMPs 2/3 and 3/9 where 64 and 69% of replacements had the same effect on the rate of hydrolysis. This discrepancy might be partially due to the fact that the results for MMP 3 were produced in a different lab than the rest of the mentioned MMPs, and different methods have been used.

Binding of larger peptides may lead to significant conformational changes that affect the binding site as follows from structural comparison of MMPs complexed with low molecular weight inhibitors and TIMPs (data not shown). The presented analysis of interaction energies may be of a limited value for MMP specificities toward larger peptides. Nevertheless, there are some data that are in accord with our results. Pairs of MMPs that were identified as similar in entire binding sites show correlations of $R^2$ in range 0.6–0.7, Fig. 2) cannot be generalized into a few simple rules that would apply to all inhibitors and substrates for all known MMPs. Rather, the results summarized in Fig. 4 and Table I and the Supplemental Material should be used to help guide synthesis of inhibitor and substrates using the data for pertinent probes and subsites. Directing the focus on optimization of ligand parts that bind in two or three subsites simultaneously could be a way to improve selectivity. For specific development projects, where ligands can have binding modes estimated from their structure, the correlations for relevant subsites and probes (Fig. 4) can be used to rationally select a set of MMPs for testing of inhibitors and small substrates to avoid redundant experiments. A correlation of inhibitor potencies toward two or more MMPs, as combined with information about subsite specificities for specific probes, can be used to infer the binding mode in terms of subsites occupied by inhibitor parts. Similarity of MMP-binding sites may provide clues for understanding of physiological roles of MMPs, at least with regard to small substrates and inhibitors. The presented data were obtained for binding sites of limited flexibility. The conclusions should be made cautiously
because they would be invalidated by larger movements of MMP structures.

REFERENCES

1. Nagase, H., and Woessner, J. F., Jr. (1997) J. Biol. Chem. 272, 14191–14194
2. Fassina, G., Ferroni, N., Brigati, C., Benelli, R., Santì, L., Noonan, D. M., and Albeni, A. (2000) Cit. Exp. Mol. Medicine 18, 111–120
3. Woessner, J. F., and Nagase, H. (2000) Matrix Metalloproteinases and TIMPs, Oxford University Press, New York
4. Massova, I., Fridman, E., and Moshayedi, S. (1997) J. Mol. Model. 3, 17–30
5. Berger, A., and Schechter, I. (1997) Philos. Trans. R. Soc. Lond.-Biol. Sci. 257, 249–264
6. Terp, G. E., Christensen, I. T., and Jorgensen, F. S. (2000) J. Biol. Chem. 275, 925–946
7. Bode, W., Fernandez, C. C., Tschese, H., Grams, F., Nagase, H., and Maskos, K. (1999) Coll. Mol. Life Sci. 55, 639–652
8. Natchus, M. G., Boukland, R. G., De, B., Natchus, M. G., Pikul, S., Almstead, N. G., De, B., Natchus, M. G., Anastasio, M. V., and McDow, D. K. (1999) Mol. Biol. 301, 513–524
9. Reiter, L. A., Mitchell, P. G., Martinelli, G. J., Lopresti-Morrow, L. L., Yocum, S., terp, G. E., Christensen, I. T., and Jorgensen, F. S. (2000) J. Mol. Biol. 304, 513–524
10. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
11. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
12. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
13. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
14. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
15. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
16. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
17. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
18. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
19. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
20. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
21. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
22. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
23. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
24. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
25. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
26. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
27. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
28. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
29. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
30. Pikul, S., McDow, D. K., Albemard, N. G., Pikul, S., Fiala, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., Pavlovsky, A. G., and Dunaway, C. M., and Moshayedi, S. (1999) J. Mol. Biol. 304, 513–524
