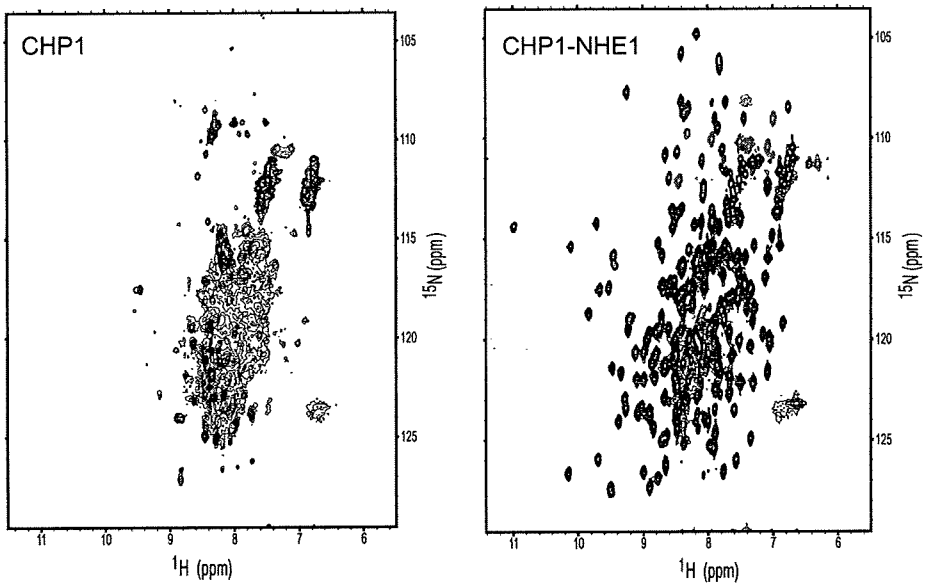

FIGURE 2. **^{1}H-^{15}N HSQC spectra of ^{15}N-labeled NHE1-free CHP1 (*left image*) and ^{15}N-labeled NHE1-(503–545)-CHP1 complex (*right image*).** These spectra were obtained with 0.3 mM samples at pH 6.9 and 37 °C recorded on the AVANCE 500.

TABLE 1

Structural statistics for NHE1-CHP1

These statistics represent an ensemble comprising 20 of the lowest energy structures obtained from 150 starting structures. Structure calculations were performed using XPLOR-NIH version 2.9.6.

Total number of distance constraints	4242
Long range ($\|i-j\|>4$)	589 (inter: 134)
Middle range ($\|i-j\|=2,3,4$)	874
Short range ($\|i-j\|=1$)	1038
Intraresidue	1521
Hydrogen bond constraints (including Ca^{2+} coordination restraints)	110 × 2
Dihedral constraints	
ϕ, φ	105, 105
$\chi 1$	17
r.m.s. deviation from experimental constraints[a]	
Distance (Å)	$0.0288 \pm 7 \times 10^{-4}$
Angle (°)	$0.44 \pm 3 \times 10^{-2}$
r.m.s. deviation from idealized covalent geometry	
Bonds (Å)	$0.00248 \pm 6 \times 10^{-5}$
Angles (°)	$0.360 \pm 6 \times 10^{-3}$
Impropers (°)	$0.31 \pm 1 \times 10^{-2}$
XPLOR energy terms (kcal/mol)[b]	
E_{bond}	23 ± 1
E_{angle}	138 ± 5
E_{imp}	28 ± 2
$E_{vdw(LJ)}$	$-6.6 \times 10^{2} \pm 0.2 \times 10^{2}$
PROCHECK Ramachandran plot (185–254)	
Residues in most favored regions (%)	78.6
Residues in additional allowed regions (%)	18.1
Residues in generously allowed regions (%)	2.9
Residues in disallowed regions (%)	0.4
r.m.s. deviation of mean structure derived from 30 calculated structures	
Backbone (10–92, 108–192, 516–538) (Å)	0.53
All heavy (10–92, 108–192, 516–538) (Å)	1.15

[a] None of these structures exhibited distance violations >0.5 or dihedral angle violations >5°.

[b] $E_{vdw(LJ)}$ represents the Lennard-Jones energy of the XPLOR energy terms.

complexed with NHE1 is well dispersed with favorable line shapes (Fig. 2), suggesting that the complex essentially adopts an ordered monomeric structure in solution.

Our target complex was ~27 kDa in size, assuming a 1:1 complex of CHP1 (22 kDa) and NHE1 (5 kDa). This represented a relatively large molecular weight in terms of conventional NMR studies. Consequently utilization of triple labeling (^{2}H, ^{13}C, and ^{15}N) and the recently developed computational methodology, CANDID, was extremely helpful in the structure determination. Sequential backbone assignments and most side-chain assignments were obtained from a 60% ^{2}H/^{15}N/^{13}C-labeled sample using standard triple resonance experiments. Missing ^{1}H resonances were supplemented using heteronuclear three-dimensional NOESY experiments with ^{15}N/^{13}C-labeled samples. Resonance assignments of methyl groups were carefully confirmed in a stereospecific manner using two-dimensional Constant Time HSQC spectra recorded for a 15% randomly enriched ^{13}C sample. Methyl groups for 18 of 24 leucines and nine of 11 valines were stereospecifically assigned. Aromatic ring proton assignments, essential for delineating hydrophobic core and protein-protein interactions, were obtained using two-dimensional TOCSY, two-dimensional NOESY, two-dimensional HCCH(arom) TOCSY, and three-dimensional ^{13}C(arom)-edited ^{15}N-separated NOESY experiments.

NMR spectra including the three-dimensional ^{13}C-edited NOESY spectrum used to monitor inter/intramolecular ^{1}H-^{1}H NOEs were of adequate quality to pursue a structural determination of the NHE1-CHP1 complex. Use of partial deuteration and the almost complete resonance assignment of methyl groups forming the hydrophobic core facilitated an initial determination of the overall protein fold. A high resolution structure was subsequently obtained using CANDID for automated assignments, which included the use of ambiguous NOEs from ^{15}N- and ^{13}C-edited NOESY experiments recorded for ^{15}N- or ^{15}N/^{13}C-labeled protein samples. An iterative approach was used for assigning NOEs in addition to the manually assigned unambiguous NOEs. The solution structure of the CHP1-NHE1 complex was determined from a total of over 4000 NMR-derived restraints, including 134 intermolecular distance restraints (Table 1). The ensemble of 20 structures in excellent agreement with a large body of experimental data were well defined (Fig. 3*A*). The r.m.s. deviations of backbone and heavy atoms over residues 518–537 of NHE1 and residues 10–92 and 108–193 of CHP1 were 0.53 and 1.15 Å, respectively. Of the NMR structures determined, the one with the smallest total energy was selected as representative for further discussion. The complex is predominantly α-helical, and the CHP1 helices constitute a cleft. A helix of the cytoplasmic region of NHE1 associates with CHP1 in 1:1 stoichiometry via the cleft (Fig. 3, *A* and *B*).

Structure of NHE1—NHE1 forms a five-turn amphipathic helix composed of residues 518–537. Orientation of the NHE1 helix is well defined relative to CHP1, consistent with the large

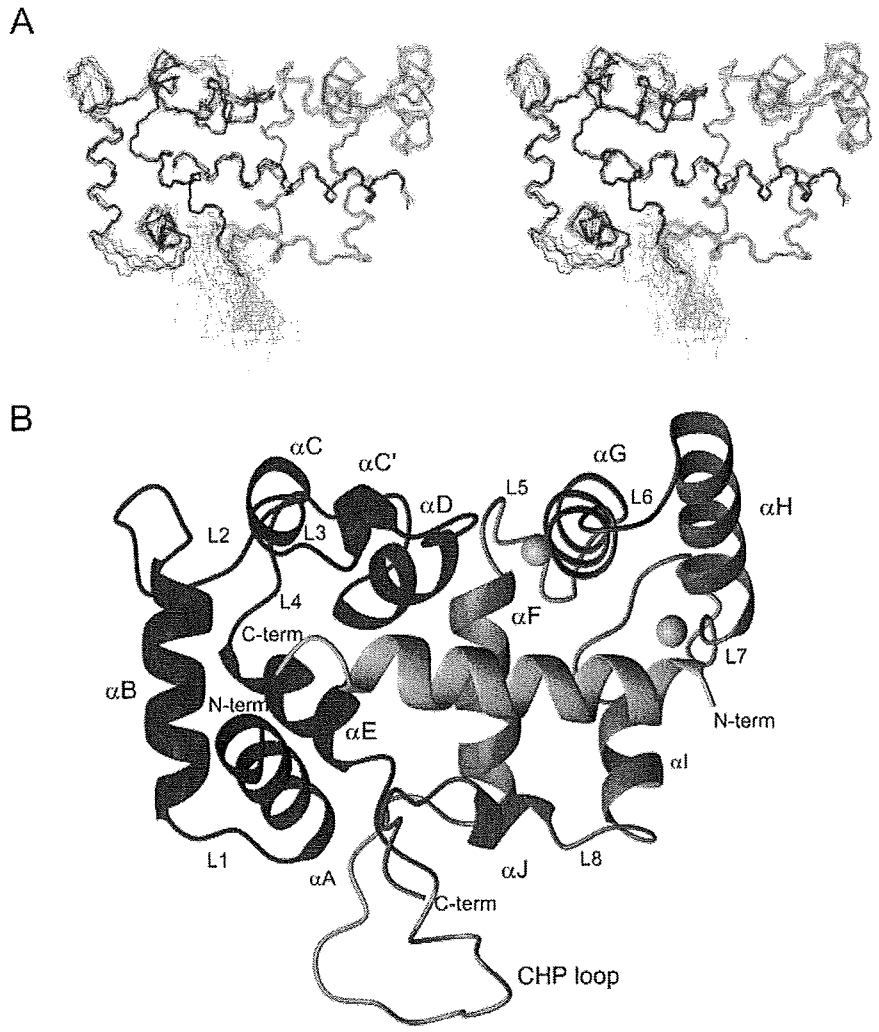

FIGURE 3. **Solution structure of the NHE1-CHP1 complex.** *A*, stereoview of the backbone superpositions of the final 20 simulated annealing structures of the NHE1-CHP1 complex. *B*, ribbon drawing of the representative NHE1-CHP1 structure complex. *A* and *B*, residues 517–538 of NHE1 and 10–192 of CHP1 are shown. The N- and C-terminal domains of CHP1 are colored in *blue* and *magenta*, respectively, and the CHP loop is colored in *gray*. NHE1 is shown in *green*. Ca^{2+} ions are shown by *gold spheres*.

number of intermolecular NOEs detected between CHP1 and NHE1 (Table 1 and Fig. 3*A*). The N-terminal half of the helix (residues 518–530) binds to the C-terminal domain of CHP1, and the C-terminal half of the helix (residues 531–537) binds to the N-terminal domain of CHP1 (Fig. 4*A*). Side-chain conformations of the helix are also well defined particularly for apolar residues that make extensive contacts with CHP1. For example, NMR spin-echo difference $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ experiments, which bring about $\chi 1$ rotamer information of aromatic side chain, showed that the His-523 and Phe-526 adopted *g*+ and *t* conformations, respectively. The helix exhibits amphipathic character in which the bulky hydrophobic residues Ile-518, Ile-522, His-523, Phe-526, Leu-527, Leu-530, Leu-531, Ile-534, and Ile-537 are clearly confined to one side, and hydrophilic residues are exposed at the other side (Fig. 4*B*). The hydrophobic residues form a continuous apolar surface (Fig. 4*B*). The main-chain and side-chain conformations of residues preceding and following the helix, residues 503–517 and 538–545, respectively, are poorly defined in the NMR structure because of the absence of medium and long range NOEs involving these regions. The narrow resonance linewidths, chemical shift index, and steady state $\{^1H\}$-^{15}N heteronuclear NOE suggest that these regions are unstructured in the complex.

Structure of CHP1—CHP1 is composed of 10 α-helices and a long loop folded into two globular regions representing the N- and C-terminal domains (Fig. 3*B*). The secondary structure consists of αA (residues 11–22), αB (residues 26–37), αC (residues 48–51), αD (residues 64–70), αE (residues 80–88), αF (residues 111–122), αG (residues 132–143), αH (residues 149–162), αI (residues 174–180), and αJ (residues 185–188) (Figs. 1*B* and 3*B*). The N-terminal domain consists of ancestral EF-hands, EF-1 and EF-2, that do not bind calcium under physiological conditions. The EF-1 hand includes helix αB, loop L2, and helix αC followed by loop L3 to the second EF-hand that includes helix αD, loop L4, and helix αE (Figs. 1*B* and 3*B*). A long loop region consisting of residues 93–110 connects the N- and C-terminal domains. Because this characteristic long insertion is not found in calcineurin B (Figs. 1*B* and 3, *A* and *B*), we refer to this long loop as the CHP loop. The absence of medium and long range NOEs, a chemical shift index, and $\{^1H\}$-^{15}N heteronuclear NOE value indicate that this region is flexible in solution. The first EF-hand in the C-terminal domain includes helix αF, loop L5, and helix αG followed by loop L6 and the second EF-hand that consists of helix αH, loop L7, and helix αI (Figs. 1*B* and 3*B*).

The four CHP1 EF-hands form a deep hydrophobic pocket, which constitutes the interaction surface for the NHE1 amphipathic α-helix. CHP1 binds to the apolar side of NHE1 with the four EF-hands through a side-by-side manner (Fig. 4*A*). This contrasts with the well known canonical CaM-target binding mode that represents a wrap-around manner in which two pairs of EF-hands bind to the target IQ motif helix on opposite sides to each other (43, 44).

Although there is modest sequence similarity between CHP1 and CaM, it should be noted that the latter interacts with a large number of proteins with various interaction modes including canonical 1:1 binding and non-canonical 1:1, 1:2, and 2:2 binding (43, 44). It has been suggested that the observed binding versatility of CaM could be derived from the variable positioning of the two domains, linked by a flexible linker, that can

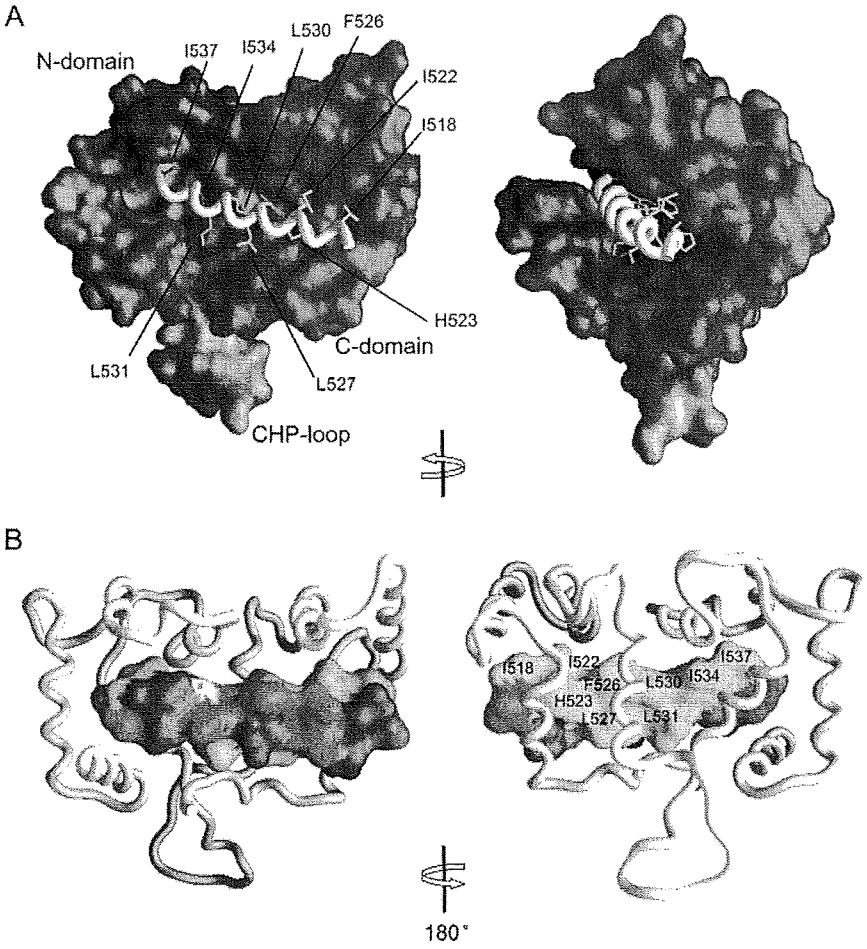

FIGURE 4. **Molecular surface of the NHE1-CHP1 complex.** *A*, molecular surface of CHP1 and backbone *tube* representation of NHE1 with *yellow stick* that shows the hydrophobic side chains. The N- and C-terminal domains of CHP1 are colored in *blue* and *magenta*, respectively. *B*, molecular surface of NHE1 and backbone *tube* representation of CHP1. Hydrophobic, acidic, and polar residues are colored in *yellow*, *red*, and *blue*, respectively.

accommodate different targets. Although target multiplicity has been reported in the case of CHP1, it seems to possess more limited binding modes than CaM. One notable feature of CHP1 that illustrates its difference to CaM has to do with the interdomain interaction. This interaction restricts domain orientation, a phenomenon absent in CaM. In our CHP1 structure, residue Leu-62 of the N-terminal domain participates in a hydrophobic interaction with residues Val-138 and Met-141 of the C-terminal domain, and residue Ala-69 of the N-terminal domain interacts with residue Leu-122 of the C-terminal domain. A side-by-side interaction mediated by four EF-hands has also been reported to take place with voltage-gated potassium channel (Kv)-interacting protein (KChIP) (45) and with CNA-CNB (46, 47) (Fig. 5*A*).

The atomic r.m.s. difference of well fitted parts between CHP1 complexed with NHE1 and CNB complexed with CNA (Protein Data Bank code 1AUI) is 2.7 Å, indicating that the fit is not very good, although the topology is identical with a high Z-score of 11.8 from a distance matrix alignment (DALI) search. The r.m.s. deviation value improves to 1.7 Å when only the N-terminal domains are superimposed, and it is 1.8 Å when only the C-terminal domains are superimposed. This indicates that the higher r.m.s. deviation for both domains originates from an interdomain swiveling between NHE1-CHP1 and CNA-CNB, although both proteins bind cognate targets in a side-by-side manner. Similarly domain swiveling was observed between NHE1-bound and NHE-free CHP1. Swiveling of the N- and C-terminal domains could create a binding surface for cognate targets.

Comparison with Other EF-hands—It is interesting to note that both EF-1 and EF-2 adopt an open conformation in NHE1-bound CHP1 without Ca^{2+}. This is especially evident when comparing the angles between the EF-hand helices. In Fig. 5*B*, a graphical view of the interhelical angles between the first and second helices of EF-1 and EF-2 of apoCaM (a typical closed conformation), Ca^{2+}-CaM (a typical open conformation, CNB), KChIP1, NHE1-bound CHP1, and NHE1-free CHP1 is displayed using a vector geometry mapping method (48, 49). The first helices of the EF-hands are superimposed along the *z* axis, and the spatial localization of the second helices are shown as a *cylinder*. This indicates that both EF-1 and EF-2 of NHE1-bound CHP1 adopt an open conformation, whereas EF-1 and EF-2 of NHE1-free CHP1 adopt an open and semiopen conformation, respectively (Fig. 5*B* and supporting information S1). This implies that EF-1 and EF-2 adopt a constitutively open conformation. However, it should be noted that the hydrophobic cleft of the N-terminal domain of free CHP1, as revealed by the crystal structure, is plugged by additional linker residues (Leu-Ala-Ala-Ala-Leu-Glu-His) (23) derived from the expression vector, partly mimicking the NHE1 helix (Fig. 5*A*). Binding of the vector-derived linker might have facilitated adoption of the open and semiopen conformations of EF-1 and EF-2, respectively. Therefore, the possibility of a "closed to open" conformational transition of EF-1 and EF-2 remains to be evaluated.

The binding of EF-hands in an open conformation without Ca^{2+} to target molecules has been found following crystal structure investigations of a Kv-KChIP1 in which the Kv fragment is covalently linked to the C terminus of KChIP1 (45). In this case, the Kv fragment binds to EF-1 and EF-2 of the N-terminal domain of KChIP1 through hydrophobic interactions (Fig. 5*A*). Of particular note, KChIP1 forms a dimer utilizing the surface formed by the Kv peptides and helix 10 of the C-terminal domain of KChIP1 in contrast to NHE1-CHP1, which exists as a monomer (Fig. 5*C*). The interhelix angles of EF-1 and EF-2

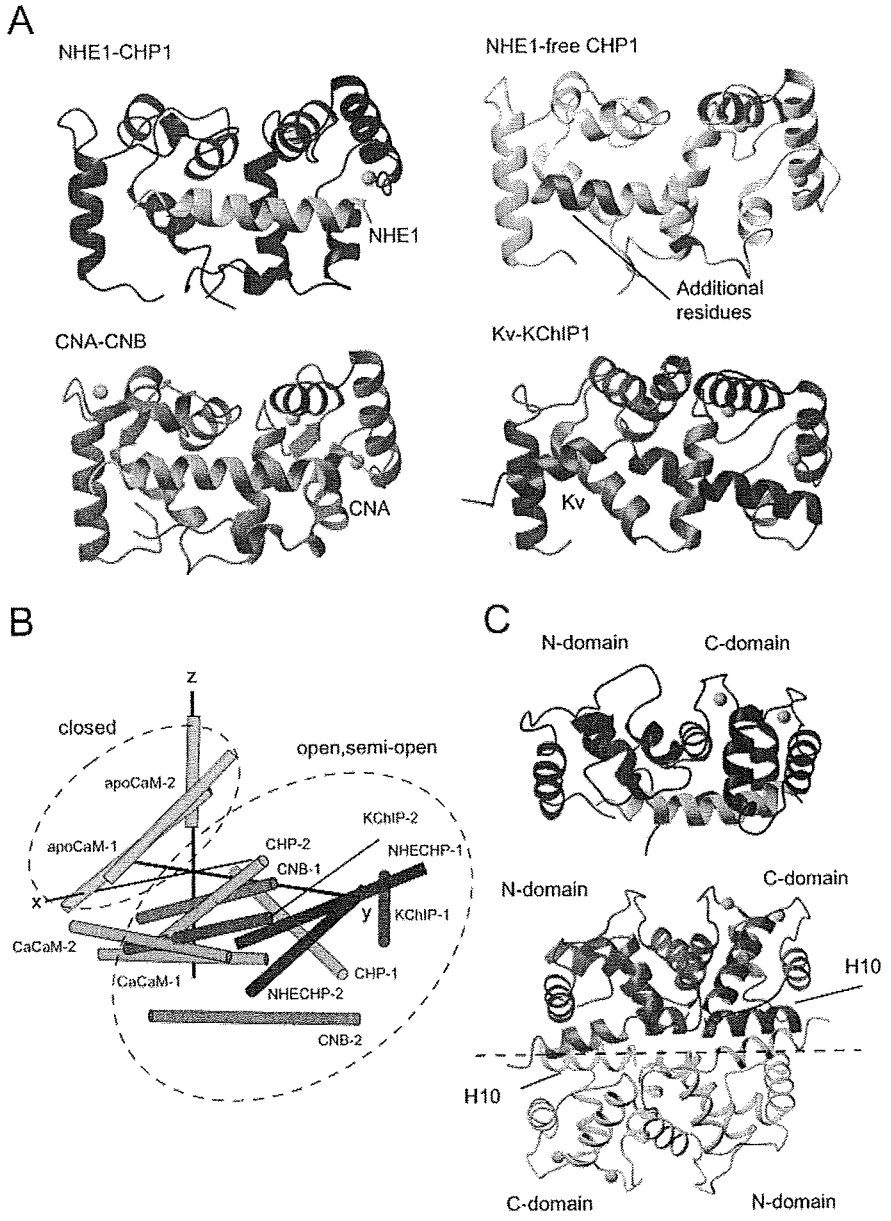

FIGURE 5. **Structure comparisons.** *A*, ribbon representation of NHE1-CHP1, rat NHE1-free CHP1 (2CT9) in which additional residues from the expression vector interacting with the N-terminal domain (LAAALEH) are depicted in *magenta*, CNA-CNB (1AUI), and Kv-KChIP1 (1S6C). Protein Data Bank entries are shown in parentheses. *B*, vector geometry mapping of EF-1 and EF-2 apocalmodulin, Ca^{2+}-calmodulin, calcineurin B, Kv-KChIP1, NHE1-bound CHP1, and NHE1-free CHP1 are denoted as *"apoCaM," "CaCaM," "CNB," "KChIP," "NHECHP,"* and *"CHP,"* respectively. *Hyphenated numbers 1* and *2* denote EF-1 and EF-2, respectively, for each protein. *C*, comparison of the KChIP1 dimer and CHP1 structures colored in *brown* and *blue*, respectively. Bound Kv and NHE1 are colored in *pink* and *green*, respectively. The dimer interface of KChIP1 is shown as a *dashed line*, and one molecule of the dimer is shown in *light colors*. KChIP1 and CHP1 are depicted in the same orientation. In *A* and *C*, the N- and C-terminal regions, αA, and the CHP loop of CHP1 are not shown in the comparison for reasons of clarity.

are 82° and 72°, respectively (supporting information S1), which are slightly larger than those of CHP1. A dimeric interaction in addition to binding of the Kv fragment may contribute to broadening of the cleft constituted by EF-1 and EF-2.

CHP1 shares 18% sequence identity with KChIP1, and the folding topology is almost identical (Fig. 5*A*). However, the target recognition mechanism differs from that of KChIP1 as judged from the determined structures. KChIP1 belongs to the extensively studied NCS1 family, which act as important regulators of various functions among certain higher eukaryotes (50). We propose that the NHE1-CHP1 interaction represents a novel binding mode utilized throughout the four-EF-hand proteins, which constitute a distinct subfamily to the NCS1 family. Furthermore detailed comparison of the binding mode of CHP1 and CNB will be presented below.

The CHP1-NHE1 Interface—The protein-protein interface consists of an extensive hydrophobic concave CHP1 undersurface and an apolar NHE1 surface. The concave undersurface spans both the N- and C-terminal domains of CHP1. The total surface area buried at the interface is 2338 $Å^2$, slightly smaller than the value of 2625 $Å^2$ for the interface between CNA and CNB (Protein Data Bank code 1AUI). Approximately more than 90% of the total buried surface area between NHE1 and CHP1 is hydrophobic, similar to the complexed structure of CNA and CNB.

The interface formed between NHE1 and CHP1 includes methyl-containing and aromatic hydrophobic residues. The residues of the N-terminal domain of CHP1, Ala-69, Phe-90, Ile-66, Thr-86, Leu-87, Phe-35, and Leu-54, comprise a hydrophobic cleft that interacts with the apolar surface formed by the side chains of NHE1 residues Leu-530, Ile-534, and Ile-537. Additionally the side chain of NHE1 Leu-531 expands the hydrophobic area interacting with CHP1 Phe-90 (Fig. 6*A*). The C-terminal domain residues of CHP1, Ala-163, Thr-159, Leu-139, Leu-135, Tyr-122, Phe-176, Ile-171, Phe-117, Ala-118, and Leu-121, constitute a continuous hydrophobic cleft that interacts with the N-terminal portion of NHE1 helix residues Ile-518, Ile-522, and Phe-526, which protrude at the apolar side. NHE1 residues Leu-527 and His-523 make additional hydrophobic contact with CHP1 residues Tyr-122, Phe-176, and Val-185 (Fig. 6*A*). These interface-forming hydrophobic residues are well conserved within NHE and CHP isoforms, reflecting the importance of these interactions (Fig. 1*B*). Furthermore the acidic side chain of Asp-528 at the center of NHE1 helix seems to form a salt bridge with the basic side

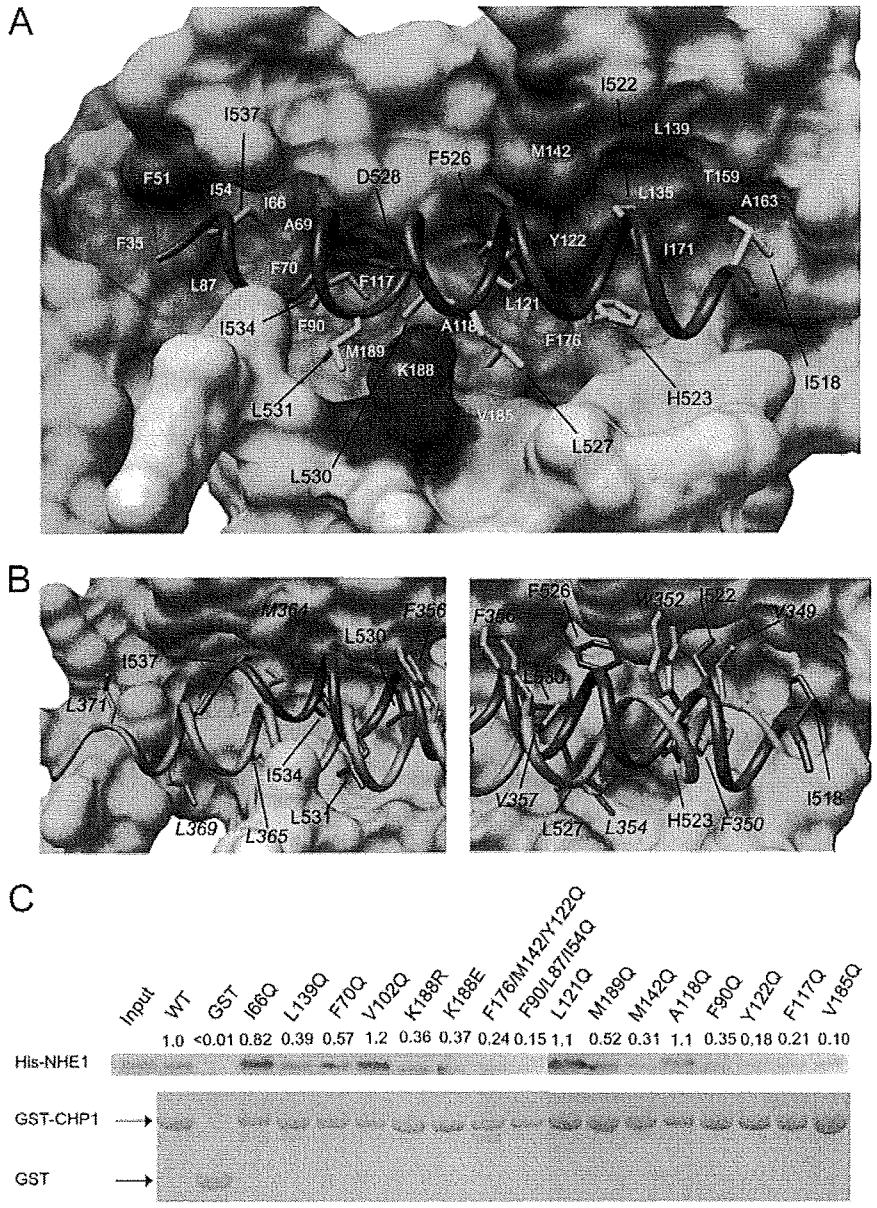

FIGURE 6. **Noncovalent interactions at the NHE1-CHP1 interface and contribution to the overall stability of the complex.** *A*, molecular surface of CHP1 and backbone *tube* representation of NHE1 with *yellow* and *red sticks* that show hydrophobic and acidic side chains, respectively. Hydrophobic and basic surface residues of CHP1 are colored in *orange* and *blue*, respectively. *B*, superposition of NHE1-CHP1 and CAN-CNB (Protein Data Bank code 1AUI). NHE1 and CNA are shown in *green* and *magenta*, respectively, and the molecular surface of only CHP1 is depicted for clarity. Complex-forming hydrophobic residues of NHE1 and CNA are shown as *sticks* and are labeled. CNA residues are represented in *italics*. The *left-hand image* and the *right-hand image* are close-up views of the N- and C-terminal domains. *C*, binding of His-NHE1 peptide to GST-CHP1 mutants in an *in vitro* pulldown assay. Bound peptide was separated via SDS-PAGE, blotted to a membrane, and then visualized using Ni-NTA-conjugated alkaline phosphatase (*top*). The amount of GST-CHP1 mutant used was estimated by Coomassie Blue staining (*bottom*) for calibration. The binding of NHE1 peptide to GST-CHP1 mutants was quantified and expressed as a ratio of binding of wild-type protein at the *top* of the gel images.

chain of CHP1 Lys-188. The presence of this interaction was observed in 10 of the final 20 NMR structures derived using the MONSTER server (51), which identifies interacting residues and assigns the nature of those interactions based on the structure. These residues are also conserved (Figs. 1, *A* and *B*, and 6*A*).

NHE1 binding orientation relative to CHP1 is probably established by surface complementarity that comprises the hydrophobic surfaces of NHE1 and CHP1. NHE1 side chains of Ile-534 and Ile-537 protrude into the shallow cleft made by the CHP1 N-terminal domain. Meanwhile NHE1 aromatic side chains of His-523 and Phe-526 protrude into the deep cleft of the CHP1 C-terminal domain (Fig. 6*A*). This deep cleft allows enough space so that they can interact with the bulky side chains of NHE1. The difference between the shapes of the domains is critical for NHE1 binding orientation.

Notably the shape of the interaction surfaces of NHE1-CHP1 and CNA-CNB differs. The spatial arrangement of NHE1 and CNA residues that contribute to form the interaction surfaces differs for each (Fig. 6*B*). Only Val-349, Phe-350, and Val-357 of CNA are located at equivalent positions to Ile-522, His-523, and Leu-530 of NHE1 unlike other residues. Cleary Trp-352 and Phe-356 of CNA that interact with the roof of the hydrophobic cleft formed by CNB are missing in NHE1. Furthermore the interactions mediated by the C-terminal stretch region of CNA, where Leu-369 and Leu-371 protrude, are absent in NHE1. On the other hand, the corresponding residues to Phe-526 and Ile-518 of NHE1 are absent in CNA.

Consequently the interaction mechanism dominated by hydrophobic interactions through amphipathic helices and four EF-hands is common in both complexes, but their shape differs. This surface complementarity in terms of a knobs-into-hole mode of interaction defines the binding specificity of these proteins, although the folding topology is identical.

Correlation with Mutagenesis Studies—The solution structure of the CHP1-NHE1 complex is essentially consistent with previous mutagenesis studies concerning NHE1. Co-immunoprecipitation experiments showed that the 4Q mutant of NHE1, in which Phe-526, Leu-527, Leu-530, and Leu-531 are substituted with glutamine, displayed no binding to CHP1, whereas the I518Q/I522Q NHE1 double mutant displayed some binding albeit with decreased affinity (3). Our structure determination revealed that Phe-526, Leu-527, Leu-530, and Leu-531 form a hydrophobic core that interacts with the center of the hydrophobic cleft of the C-terminal domain of

CHP1, whereas Ile-518 and Ile-522 interact with the rim of the hydrophobic cleft of the C-terminal domain of CHP1 (Figs. 4*B* and 6*A*). Deletion mutagenesis indicated that NHE1 (residues 510–575) retained binding affinity similar to the wild-type protein, implying that the juxtamembrane region of NHE1 comprising residues 503–509 is unimportant for CHP1 binding (3). These residues were found to be unstructured in the present study and form no direct contact with NHE1.

The binding activity of NHE1 (residues 530–656) lacking the N-terminal segment was found to be completely impaired (3). Our NHE1-CHP1 complex structure shows that an absence of residues preceding Leu-530 results in almost complete loss of interaction between NHE1 and the C-terminal domain of CHP1, whereas the many hydrophobic interactions mediated by NHE1 residues Leu-530, Leu-531, Ile-534, and Ile-537 are retained (Fig. 6*A*). This indicates that interaction of the N-terminal residues of NHE1 and the C-terminal domain of CHP1 is indispensable in maintaining the NHE1-CHP1 complex. Namely this part of the interaction plays a dominant role in NHE1-CHP1 complex formation.

CHP1 was subjected to site-directed mutagenesis in an effort to reveal detailed individual contributions of interfacial residues toward the overall stability of the NHE1-CHP1 complex. We selected three categories of residues in CHP1: 1) residues that comprise the floor of the hydrophobic pocket, Phe-70, Phe-90, Phe-117, Ala-118, Leu-121, Tyr-122, and Met-189; 2) residues forming the rim of the pocket, Ile-66, Leu-139, Met-142, and Val-185; and 3) a residue forming the salt bridge, Lys-188. A noninteracting solvent-exposed residue, Val-102, as the positive control and a triple mutation involving hydrophobic residues Phe-90/Leu-87/Leu-54 and Phe-176/Met-142/Tyr-122 as the negative control were prepared.

Dramatic effects were observed in terms of CHP1 binding to NHE1 with mutations F117Q, Y122Q, and V185Q (Fig. 6*C*). We expected the floor-forming residues located at the bottom of the cleft to be critical for the interaction, but the rim-forming residue Val-185 showed significant reduction similar to the negative control. Based on the solution structure of the NHE1-CHP1 complex, it was appeared that Phe-117, Tyr-122, and Val-185 were confined to the C-terminal domain of CHP1 and were packed against the apolar side of NHE1. This suggests that the C-terminal hydrophobic cleft represents a mutation hot spot, implying that it plays a key role in the NHE1-CHP1 interaction. This result is consistent with the NHE1 deletion result indicating that interaction between the N-terminal segment of NHE1 and the C-terminal domain of CHP1 is dominant. This represents a unique feature of the NHE1-CHP1 interaction.

Although mutation of residues possibly involved in salt bridge formation such as K188E and K188R resulted in a marked decrease in binding interaction, the effect was not strong, and the charge-reversing effect of K188E was unclear. Thus, it appears that although this salt bridge contributes to NHE1-CHP1 complex formation, it is not the main force possibly because this bond is exposed to solvent (Fig. 6*A*), and therefore solvation could weaken the strength of the interaction.

Role of CHP1—NMR investigations of NHE1 complexed with CHP1 revealed that the juxtamembrane region comprising residues 503–517 was unstructured in solution. This region is rich in basic residues with a previous study reporting that NHE1 comprising residues 506–576 bound to PIP_2 *in vitro* (13). In addition, residues 513–520 and 556–564 might represent PIP_2 binding sites (13). Accordingly it is likely that this juxtamembrane region (residues 503–517) following the last transmembrane helix (H12, residues 478–499) is incapable of forming a continuous straight helical structure into the cytoplasm due to interaction with the membrane. Rather the overall structure of NHE1 presumably turns or bends toward the cytoplasmic membrane following a PIP_2-mediated interaction. Assuming that flanking regions of the helix, residues 513–520 and 556–564, attach to the membrane (PIP_2) (13), the helix and CHP1 should be located immediately beneath the cytoplasmic membrane.

Furthermore it has been reported that NHE1 acts as a scaffold protein linked to actin filaments via ezrin-radixin-moesin proteins in addition to possessing function as an ion exchanger (52, 53). Our NMR studies revealed that the cytoplasmic helix-forming residues of NHE1 comprise residues 518–537, which showed little overlap from the previously reported ezrin-radixin-moesin binding region (512–520 and 550–565). This ensures simultaneous binding of NHE1 to CHP1 and ezrin-radixin-moesin proteins.

The juxtamembrane region of NHE1 forms a particular tertiary or quaternary structure that is mediated by interactions with the membrane (PIP_2), CHP1, and ezrin-radixin-moesin proteins. The overall structure of the juxtamembrane region might play an important role in NHE1 activity. The 90% loss in activity following CHP1 depletion might be due to disruption of the structure of the cytoplasmic region of NHE1 around the membrane. In fact, although this represents *in vitro* evidence, the amphipathic helix is disrupted in the absence of CHP1 based on the CD and NMR data (supporting information S2 and S3).

Possible Mode of Regulation—CHP1 deprivation resulted in impaired regulation of NHE1 following external stimuli, implying that CHP1 acts as regulator of NHE1 by involvement in the processing of intracellular signals derived from external stimuli. However, the regulatory mechanism remains unclear. Although it was reported that CHP1 is an *N*-myristoylated protein, CHP1 does not exert myristoyl switching in a Ca^{2+}-dependent manner under normal physiological conditions because EF-3 and EF-4 constitutively bind Ca^{2+} ions where *N*-myristoylation was not required for NHE1 binding, activation, or localization (14).

Of particular note, it was reported that CHP1 is a phosphorylated protein, although the phosphorylation sites were not determined. According to the phosphorylation prediction server NetPhos (54), the CHP1 sequence contains potential phosphorylation sites located at residues Thr-36, Ser-37, Ser-47, Ser-131, and Ser-172 (score, >0.8). Similarly the phosphorylation server Scansite (55) identified potential phosphorylation sites located at residues Thr-36, Ser-37, and Ser-172 (score, >0.5). From the determined NHE1-CHP1 structure, residues Thr-36, Ser-37, and Ser-172, predicted by both servers as potential phosphorylation sites, are locate at the terminal part of the EF-hand helix or its flanking loop where the side chains of

the aforementioned residues are exposed to the protein surface reinforcing the possibility of phosphorylation. Phosphorylation-induced conformational changes in CHP1 and the subsequent regulation of NHE1 activity are interesting areas that remain to be investigated.

Conclusion—We have determined the solution structure of the cytoplasmic region of NHE1 complexed with CHP1. Although previous biochemical analyses suggested that the hydrophobic residues of NHE1 were likely to interact with CHP1, the present study has delineated the structural basis for this interaction. The solution structure provides concrete evidence that the cytoplasmic region of NHE1 forms an amphipathic helix that interacts directly with the large concave undersurface of CHP1. This helix is disrupted in the absence of CHP1; thus the loss in activity following CHP1 depletion might be due to disruption of the structure of the juxtamembrane region of NHE1. Our structure provides a first step toward understanding the regulation of NHE1 activity. Moreover it revealed a novel target binding mechanism mediated by four EF-hands. These findings should facilitate future studies aimed at understanding the mechanism underlying recognition utilized by EF-hand proteins that are engaged in signal transduction pathways and many other molecular and cellular events. During the initial review of our manuscript, a study appeared that describes a crystal structure of the NHE1 peptide complexed with CHP2 containing Y^{3+} ions instead of Ca^{2+} (56).

Acknowledgments—We are grateful to Momoko Yoneyama, Hiroko Kinoshita, and Junko Tsukamoto of the Nara Institute of Science and Technology for technical assistance and Kokoro Hayashi for help in sample preparation.

REFERENCES

1. Orlowski, J., and Grinstein, S. (2004) *Pfluegers Arch. Eur. J. Physiol.* **447,** 549–565
2. Lin, X., and Barber, D. L. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12631–12636
3. Pang, T., Su, X., Wakabayashi, S., and Shigekawa, M. (2001) *J. Biol. Chem.* **276,** 17367–17372
4. Bertrand, B., Wakabayashi, S., Ikeda, T., Pouyssegur, J., and Shigekawa, M. (1994) *J. Biol. Chem.* **269,** 13703–13709
5. Wakabayashi, S., Bertrand, B., Ikeda, T., Pouyssegur, J., and Shigekawa M. (1994) *J. Biol. Chem.* **269,** 13710–13715
6. Dhanasekaran, N., Prasad, M. V., Wadsworth, S. J., Dermott, J. M., and van Rossum, G. (1994) *J. Biol. Chem.* **269,** 11802–11806
7. Hooley, R., Yu, C. Y., Symons, M., and Barber, D. L. (1996) *J. Biol. Chem.* **271,** 6152–6158
8. Voyno-Yasenetskaya, T., Conklin, B. R., Gilbert, R. L., Hooley, R., Bourne, H. R., and Barber, D. L. (1994) *J. Biol. Chem.* **269,** 4721–4724
9. Bianchini, L., L'Allemain, G., and Pouyssegur, J. (1997) *J. Biol. Chem.* **272,** 271–279
10. Takahashi, E., Abe, J., Gallis, B., Aebersold, R., Spring, D. J., Krebs, E. G., and Berk, B. C. (1999) *J. Biol. Chem.* **274,** 20206–20214
11. Lehoux, S., Abe, J.-i., Florian, J. A., and Berk, B. C. (2001) *J. Biol. Chem.* **276,** 15794–15800
12. Yan, W., Nehrke, K., Choi, J., and Barber, D. L. (2001) *J. Biol. Chem.* **276,** 31349–31356
13. Aharonovitz, O., Zaun, H. C., Balla, T., York, J. D., Orlowski, J., and Grinstein, S. (2000) *J. Cell Biol.* **150,** 213–224
14. Pang, T., Hisamitsu, T., Mori, H., Shigekawa, M., and Wakabayashi, S. (2004) *Biochemistry* **43,** 3628–3636
15. Barroso, M. R., Bernd, K. K., DeWitt, N. D., Chang, A., Mills, K., and Sztul, E. S. (1996) *J. Biol. Chem.* **271,** 10183–10187
16. Lin, X., Sikkink, R. A., Rusnak, F., and Barber, D. L. (1999) *J. Biol. Chem.* **274,** 36125–36131
17. Timm, S., Titus, B., Bernd, K., and Barroso, M. (1999) *Mol. Biol. Cell* **10,** 3473–3488
18. Matsumoto, M., Miyake, Y., Nagita, M., Inoue, H., Shitakubo, D., Takemoto, K., Ohtsuka, C., Murakami, H., Nakamura, N., and Kanazawa, H. (2001) *J. Biochem. (Tokyo)* **130,** 217–225
19. Nakamura, N., Miyake, Y., Matsushita, M., Tanaka, S., Inoue, H., and Kanazawa, H. (2002) *J. Biochem. (Tokyo)* **132,** 483–491
20. Pang, T., Wakabayashi, S., and Shigekawa, M. (2002) *J. Biol. Chem.* **277,** 43771–43777
21. Mailander, J., Muller-Esterl, W., and Dedio, J. (2001) *FEBS Lett.* **507,** 331–335
22. Perera, E. M., Martin, H., Seeherunvong, T., Kos, L., Hughes, I. A., Hawkins, J. R., and Berkovitz, G. D. (2001) *Endocrinology* **142,** 455–463
23. Naoe, Y., Arita, K., Hashimoto, H., Kanazawa, H., Sato, M., and Shimizu, T. (2005) *J. Biol. Chem.* **280,** 32372–32378
24. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc.* **116,** 11655–11666
25. Matsuo, H., Kupce, E., Li, H., and Wagner, G. (1996) *J. Magn. Reson. B* **111,** 194–198
26. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson. B* **103,** 203–216
27. Clowes, R. T., Boucher, W., Hardman, C. H., Domaille, P. J., and Laue, E. D. (1993) *J. Biomol. NMR* **3,** 349–354
28. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Formankay, J. D. (1993) *J. Magn. Reson. B* **101,** 333–337
29. Logan, T. M., Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1993) *J. Biomol. NMR* **3,** 225–231
30. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) *Biochemistry* **28,** 7510–7516
31. Boucher, W., Laue, E. D., Campbell-Burk, S., and Domaille, P. J. (1992) *J. Am. Chem. Soc.* **114,** 2262–2264
32. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, pp. 301–531, Academic Press, San Diego, CA
33. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3,** 185–204
34. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–279
35. Goddard, T. D., and Kneller, D. G. (1999) *SPARKY3*, University of California, San Francisco
36. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J. Mol. Biol.* **319,** 209–227
37. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J. Magn. Reson.* **160,** 65–73
38. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8,** 477–486
39. Koradi, R., Billete, M., and Wuthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55
40. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11,** 281–296
41. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J. Comput. Chem.* **25,** 1605–1612
42. Goda, N., Tenno, T., Takasu, H., Hiroaki, H., and Shirakawa, M. (2004) *Protein Sci.* **13,** 652–658
43. Hoeflich, K. P., and Ikura, M. (2002) *Cell* **108,** 739–742
44. Bhattacharya, S., Bunick, C. G., and Chazin, W. J. (2004) *Biochim. Biophys. Acta* **1742,** 69–79
45. Zhou, W., Qian, Y., Kunjilwar, K., Pfaffinger, P. J., and Choe, S. (2004) *Neuron* **41,** 573–586
46. Griffith, J. P., Kim, J. L., Kim, E. E., Sintchak, M. D., Thomson, J. A., Fitzgibbon, M. J., Fleming, M. A., Caron, P. R., Hsiao, K., and Navia, M. A. (1995) *Cell* **82,** 507–522
47. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., Gastinel, L. N., Habuka, N., Chen, X., Maldonado, F., Barker, J. E., Bacquet, R., and Villafranca, J. E. (1995) *Nature* **378,** 641–644
48. Yap, K. L., Ames, J. B., Swindells, M. B., and Ikura, M. (1999) *Proteins* **37,**

499–507
49. Yap, K. L., Ames, J. B., Swindells, M. B., and Ikura, M. (2002) *Methods Mol. Biol.* **173,** 317–324
50. Burgoyne, R. D., and Weiss, J. L. (2001) *Biochem. J.* **353,** 1–12
51. Salerno, W. J., Seaver, S. M., Armstrong, B. R., and Radhakrishnan, I. (2004) *Nucleic Acids Res.* **32,** W566–W568
52. Baumgartner, M., Patel, H., and Barber, D. L. (2004) *Am. J. Physiol.* **287,** C844–C850
53. Denker, S. P., Huang, D. C., Orlowski, J., Furthmayr, H., and Barber, D. L. (2000) *Mol. Cell* **6,** 1425–1436
54. Blom, N., Gammeltoft, S., and Brunak, S. (1999) *J. Mol. Biol.* **294,** 1351–1362
55. Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003) *Nucleic Acids Res.* **31,** 3635–3641
56. Ammar, Y. B., Takeda, S., Hisamitsu, T., Mori, H., and Wakabayashi, S. (2006) *EMBO J.* **25,** 2315–2325

A Simple Method to Improve the Odds in Finding 'Lead-Like' Compounds from Chemical Libraries

Kouhei HORIO,[a] Hajime MUTA,[a] Junichi GOTO,[b] and Noriaki HIRAYAMA*,[a]

[a] *Tokai University School of Medicine; Boseidai, Isehara, Kanagawa 259–1193, Japan: and* [b] *Computational Science Department, Science & Technology Systems Division, Ryoka Systems Inc.; 1–28–38 Shinkawa, Chuo-ku, Tokyo 104–0033, Japan.* Received November 20, 2006; accepted April 20, 2007; published online April 25, 2007

A simple method of virtual screening is proposed. This method uses only chemical characters calculated from two dimensional chemical structures. Local and global chemical characters are represented by molecular fingerprint and trait, respectively. The trait is a newly introduced concept in this paper and it is expressed by a set of two dimensional (2D) molecular descriptors. In this study, both the molecular fingerprint and the trait were used to represent drug-likeness of a group of molecules with a particular pharmacological activity. To learn about the molecular fingerprint and trait specific to a particular group of drugs, we used a database of drugs that are clinically used in Japan now. The molecular fingerprint and trait trained on these real drugs were used to predict drug-likeness of molecules in other chemical databases. In these chemical databases, an appreciable number of compounds that show the relevant pharmacological activity are contained. Some of these compounds are drugs clinically used abroad, but not in Japan. The prediction rate was judged by an enrichment factor. Despite the simplicity of the methodology, practical results were obtained. In the case of β-adrenergic blockers, the enrichment factor of 66 was attained and nearly 57% of active molecules in the chemical databases were successfully covered.

Key words virtual screening; drug-likeness; molecular descriptor; molecular fingerprint; trait

A large number of new technologies have been introduced into pharmaceutical research in recent years to provide information about potential drug targets and their disease association. In most drug discovery projects, however, we are still in the situation where one does not have any specific target in mind at the time when a set of molecules for screening is necessary. In this situation, we often use small number of active compounds as clues to discover better drug candidates. Since quite a large number of chemicals are commercially available, the discovery process usually starts from screening drug candidates in these chemical databases. In pharmaceutical companies, the proprietary compounds are also added to the target databases to be screened. A tremendous number of compounds is usually contained in such databases. Therefore, it is usually not realistic that even modern high-throughput screening technique can actually handle all of the compounds. If it is possible to reduce the number of compounds to be experimentally tested, it can make the entire process of searching cost effective and fast. Virtual screening can serve as a tool to suggest compounds to be tested experimentally. In this sense, virtual screening is a computational counterpart to high-throughput screening. As virtual screening is an entirely computer-based method, it can search a very large chemical space. Virtual screening is becoming an important screening technique because it is also very cost effective.

In this paper, a simple method of virtual screening is proposed. It is usually difficult to identify the specific conformation of a drug that is responsible to the biological activity. Accordingly, in many drug discovery projects, we have no alternate but to use only two dimensional chemical structures. Since chemical characters of drug molecules determine their drug-likeness, this method uses only chemical characters calculated from the two-dimensional chemical structures of the molecules. Local and global chemical characters are represented by molecular fingerprint and trait, respectively. The concept of molecular fingerprint is relatively well-established in the field of cheminformatics. As a molecular fingerprint is expressed by substructural information, the molecular fingerprint represents a local chemical character of a molecule. The trait, however, is newly introduced concept in this study and it can represent a global character of a set of molecules. A trait of a particular group of molecules is expressed by a combination of various two dimensional (2D) molecular descriptors. The 2D molecular descriptors only use the atoms and connection information of the molecule. The trait represents common global characters of these molecules. A drug-likeness profile of a set of drugs with a particular pharmacological activity can be expressed by a combination of the molecular fingerprint and the trait. To learn about drug-likeness of drugs with a specific pharmacological activity, we used a database of drugs that are clinically used in Japan (DCUJ) now. The number of independent drugs in DCUJ was 1170 when we undertook this study. Namely, DCUJ was used as a training dataset. The drug-likeness profile trained on these real drugs was employed to predict the specific drug-likeness of compounds in other chemical databases. In these chemical databases which consist of thousands of compounds, an appreciable number of compounds that show the relevant pharmacological activity are contained. Some of these compounds are drugs clinically used abroad, but not in Japan. The prediction rate was judged by an enrichment factor. Despite the simplicity of the methodology, reasonably practical results have been successfully obtained.

Experimental

Database About 1200 independent drug molecules are contained in DCAJ and these molecules were used to make training datasets. Drugs belonging to four categories *i.e.* anxiolytic agents, β-adrenergic blockers, ACE inhibitors and antibiotics were used to make the training dataset for each drug category. There are several chemical databases that contain many drugs and drug-like compounds. These databases can be used as test datasets for validation. Merck Index (Merck)[1] and Comprehensive Medicinal Chemistry (CMC)[2] contain many drug and drug-like compounds. The information of

* To whom correspondence should be addressed. e-mail: hirayama@is.icc.u-tokai.ac.jp

investigative new drugs is also useful to consider the drug-likeness of compounds. In this study a database of the investigative new drugs under clinical trials in Japan (IND)[3] is taken into account. We merged these three databases and removed the compounds contained in DCUJ and other redundant compounds to construct a test dataset consisting of 8277 unique compounds. The test dataset was used to evaluate the predictive ability of the method mentioned below.

Molecular Fingerprint Binary forms of chemical structures based on 2D chemical representation are called molecular fingerprints. Several methodologies exist for chemical binary representations. Molecular Design Limited (MDL) created a key based molecular fingerprint. This molecular fingerprint uses a pre-defined set of definitions and creates molecular fingerprints based on pattern matching of the structure to the defined "key" set.[4] This key based approach relies on the definitions to encapsulate the molecular descriptions. The keys were originally developed for the purpose of database substructure searching. Each "key" describes a small substructure consisting of about one to ten non-hydrogen atoms. These substructures can be used to characterize molecules. Molecular fingerprints also can be used to filter out compounds that do no meet certain criteria. In this study, the bit packed version of the 166 public MDL MACCS structural keys was used to represent molecular fingerprints.

Trait Molecular fingerprints only represent local molecular characters. Global characters of drugs are also highly important to determine drug-likeness of molecules. We have introduced a new concept of trait to express common global characters of a set of drugs that possess a similar pharmacological activity. To represent the trait of drugs of a particular pharmacological category, the distribution patterns of 2D descriptors were used. In this study, 124 descriptors as shown in Table 1 were used.

Most of the descriptors in Table 1 are described in ref. 5 and implemented in the software system MOE.[6] The asterisked descriptors are newly introduced for this study. Brief explanations for these descriptors are as follows:

(1) contents of atoms

a_wY: the weight percent of the atom type Y in a molecule. The *sp*, sp^2 and sp^3 hybridization states of atoms are expressed by sp, sp2 and sp3, respectively. Hali, Haro and Hact mean hydrogen atoms bonded to aliphatic carbon atoms, aromatic carbon atoms and other types of atoms, respectively.

(2) molecular dimension

'd' and 'r' are the largest and smallest values in the distance matrix, respectively. d/r is used to represent the relative shape of the molecule.

(3) bonds

b_xxxxR: number of bonds of bond type xxxx (including implicit hydrogens) divided by the number of all bonds (including implicit hydrogens). "ar" denotes aromatic bond.

b_count_mw: number of bonds (including implicit hydrogens) divided by molecular weight

b_heavy_mw: number of bonds between heavy atoms divided by molecular weight

(4) connectivity and shape index

chiN_mw: atomic connectivity index (order N) divided by molecular weight

chiNv_mw: atomic valence connectivity index (order N) divided by molecular weight.

chiN_Cmw: carbon connectivity index (order N) divided by molecular weight.

chiNv_Cmw: carbon valence connectivity index (order N) divided by molecular weight.

It is not easy to consider the correlations between the molecular descriptors and find the independent descriptors. For instance, molecular weight and volume are generally highly correlated, but it is obvious that there are not always correlated. Therefore in this study all molecular descriptors are treated independently. The ranges of the values of 124 molecular descriptors were calculated for a set of drugs of a particular category. Then the ranges that cover a certain percentage of the drugs were determined. A trait of a particular group of drugs is expressed by a set of these ranges of the 124 molecular descriptors.

Traits of several drug categories expressed by a small set of descriptors are shown in Table 2. In these examples, 14 typical molecular descriptors are used to represent global characters, *i.e.* traits. The range for each descriptor was determined in order to cover 80% of drugs with the same pharmacological activity incorporated in DCUJ. The range is characteristic of each drug category. The ranges of the descriptors observed for anxiolytic drugs are relatively narrow. For antibiotics, however, the ranges are appreciably wide. The ranges of the descriptors that cover 80% of all drugs ('drugs' in Table 2) in DCUJ are also noteworthy. They, all together, represent a drug-likeness profile of drugs in general. By use of 124 descriptors instead of 14, we can more appropriately define the trait of each drug category. The ranges of all 124 descriptors that cover 80% of all drugs in DCUJ are given Supplemental Table S1.

Table 1. 124 Descriptors Used to Express the Trait of Molecule

contents of atoms
a_wBr,* a_wCl,* a_wF,* a_wI,* a_wP,* a_wS,* a_wN,* a_wNsp,* a_wNsp2,* a_wNsp3,* a_wO,* a_wOsp2,* a_wOsp3,* a_wH,* a_wHali,* a_wHaro,* a_wHact,* a_wC,* a_wCsp,* a_wCsp2,* a_wCsp3*
grobal properties
weight, vdw_vol, density, d/r,* vdw_area, RingN (Oprea), SlogP, TPSA, SMR, apol, bpol, Fcharge
bonds
b_singleR,* b_doubleR,* b_tripleR,* b_arR,* b_count_mw,* b_heavy_mw*
pharmacophore features
a_acc, a_acid, a_base, a_don, a_hyd
information content
a_IC, a_ICM
bond rotation
b_1rotN, b_rigidN (Oprea), b_rotN, b_1rotR, b_rotR
connectivity and shape indices
chi0_mw,* chi1_mw,* chi0v_mw,* chi1v_mw,* chi0_C_mw,* chi1_C_mw,* chi0v_C_mw,* chi1v_C_mw,* KierA1, KierA2, KierA3, KierFlex, Zagreb
vertex
VAdjEq, VAdjMa.
adjacency and distance matrix descriptors
VDistEq, VdistMa, balabanJ
subdivided surface areas
SlogP_VSA0, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SMR_VSA0, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7
surface areas
vsa_acc, vsa_acid, vsa_base, vsa_don, vsa_hyd, vsa_other, vsa_pol
partial charge descriptors
PEOE_PC−, PEOE_PC+, PEOE_RPC−, PEOE_RPC+, PEOE_VSA_POS. PEOE_VSA_FPOS, PEOE_VSA_PPOS, PEOE_VSA_FPPOS, PEOE_VSA_NEG, PEOE_VSA_FNEG, PEOE_VSA_PNEG, PEOE_VSA_FPNEG, PEOE_VSA_POL, PEOE_VSA_FPOL, PEOE_VSA_HYD, PEOE_VSA_FHYD, PEOE_VSA+0, PEOE_VSA+1, PEOE_VSA+2, PEOE_VSA+3, PEOE_VSA+4, PEOE_VSA+5, PEOE_VSA+6, PEOE_VSA−0, PEOE_VSA−1, PEOE_VSA−2, PEOE_VSA−3, PEOE_VSA−4, PEOE_VSA−5, PEOE_VSA−6

Asterisked descriptors are made for the present study.

Drug-Likeness and Enrichment Enrichment procedure by use of molecular fingerprints was undertaken by the molecular fingerprint model function implemented in MOE. Molecular fingerprints calculated for drugs in the training dataset were used as the reference molecular fingerprints. The molecular fingerprints of drugs in the training dataset were compared to the molecular fingerprints of compounds in the test dataset. The similarity between all drugs in the training dataset and a compound in the test dataset was judged by the concordance between their molecular fingerprints. The compounds with molecular fingerprints that exceeded the specified similarity thresholds were regarded as compounds that should possess the relevant biological activity. The Tanimoto index[7] was used as a similarity metric and 80% was used as a specified similarity threshold in the present study. 80% was a good threshold that covers highly similar compounds.

Ranges of the values of 124 descriptors calculated for a group of compounds can characterize the common features of the group. Therefore the ranges that cover 80% of drugs with a particular pharmacological activity should well represent the trait of this family of drugs. The trait was used to judge a propensity of a compound for having the specific pharmacological activity. If a descriptor of a compound in a test dataset took a value within the range defined by the trait, a 1(hit) was assigned. If the value was out of the range, a 0 was assigned. The number of hits divided by 124 takes a continuous value from 0 to 1. The drug probability based on this value was used to find drug-like compounds. In this study the compounds that are ranked top 2% are considered as drug-like compounds that should possess the rele-

vant activity. The number of compounds ranked top 2% was reasonable from the practical point of view in this study. The percentage can be varied according to a required enrichment level.

By use of the criteria of drug-likeness described above, the most probable drug-like compounds in the test dataset were clustered.

Enrichment Factor Virtual screening is used to suggest compounds in a dataset, which are ideally of high affinity to a certain receptor. For validation of the suggested compound one has to check the outcome against a sample that is chosen by random. The enrichment factor describes how much higher the proportion of hits (*i.e.* active compounds) is in any given sample of entities compared with the randomly picked sample. When performing virtual screening, the ability of achieving enriched samples is much prized than the ability of predicting single affinities with high accuracy. The formula for the enrichment factor (EF) is usually expressed as follows[8]:

$$EF = V_A/(V(L_A/L)) = LV_A/VL_A$$

Here L and L_A mean the total number of compounds in a library and the number of experimentally verified bioactive compounds in the library, respectively. V and V_A mean the number of compounds predicted as active and the number of known bioactive compounds contained in V, respectively. An enrichment factor of 1 is equally good as random selection.

Software Two programs coded by use of the scientific vector language (SVL) of the software system MOE were made for the present study. The first program DRF is used to analyze the distributions of 2D descriptors and define the ranges of 2D descriptors that cover the sufficient number of compounds. A typical distribution of a descriptor SlogP is shown in Fig. 1. In this figure the range of SlogP that covers 80% of the training dataset is shown. As the distribution curve is a smooth and monomodal one, SlogP is considered as a good descriptor to be used as an identifier of drug-likeness. The second program DRFF calculates the drug probabilities of compounds in the test dataset by use of the ranges of descriptors that were determined by DRF. The source codes of these two programs written by SVL are available from Ryoka Systems Inc.[9] free of charge.

Results and Discussion

The drug-likeness trained by the training dataset was applied in the prediction of active molecules in the test dataset to evaluate the predictive power of this procedure. The summary of the results are given in Table 3. The number of anxiolytic compounds in the test dataset is 274. The trait of anxiolytic compounds was calculated based on the anxiolytic compounds contained in DCUJ. The drug probability as an anxiolytic agent was calculated using this trait for each compound in the test dataset. The top 2% of compounds with the highest drug probability was selected as active compounds (TR). The number of top 2% slightly varies depending on the values of drug probability. In the case of anxiolytic compounds, it was 198. Out of 198 compounds, the number of compounds with actual anxiolytic activity (TR(active)) was 47. Therefore the enrichment factor of anxiolytic compounds based on trait (EF(TR)) was 6.8. The molecular fingerprints of anxiolytic compounds were trained by use of DCUJ. The compounds with the similarity metrics greater than 80% were screened from the test dataset and considered as hits. Out of 123 hits (FP), 46 compounds were experimentally verified active compounds (FP(active)). The enrichment factor attained by use of molecular fingerprint (EF(FP)) was 10.6. The union (FP∪TR) and product (FP∩TR) of the screened compounds by trait and molecular fingerprint show complementarities of these two filters. The number of compounds selected by the product of FP and TR (FP∩TR) is 34 and the number of known active compounds (FP∩TR(active)) contained in it was 26. Although the number of active compounds screened is just 9.0% of the total active compounds, the rate of false positive is quite low and an enrichment factor (EF(FP∩TR)) of 21.8 was attained. If the union of the compounds screened by both of the filters is considered (FP∪TR), the number of hits would increase (67) at the

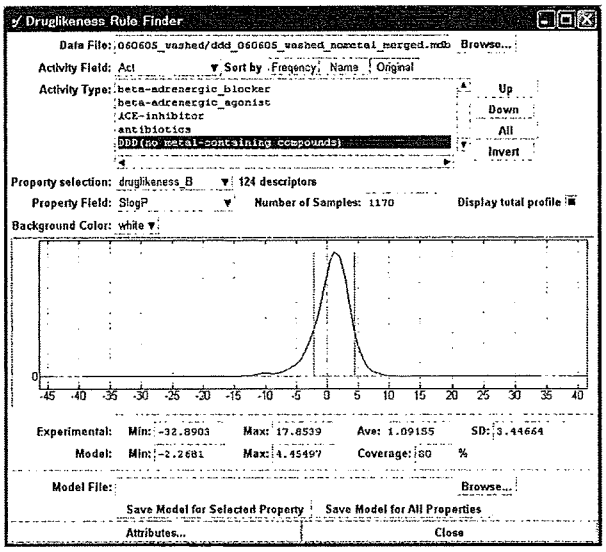

Fig. 1. The Distribution of the SlogP Values of the Drugs Clinically Used in Japan Now

Under the peak between the two vertical lines, 80% of the drugs are included. The horizontal and vertical axes denote the SlogP value and the population of the drugs, respectively.

Table 2. Traits of Drugs Expressed by the Ranges of 14 Typical Molecular Descriptors

	Anxiolytic agent		β-Adrenergic blocker		ACE-Inhibitor		Antibiotics		Drugs[a]	
Weight	301	377	249	373	368	439	323	751	194	566
SlogP	1.27	4.37	−0.09	1.86	−0.93	0.47	−6.45	1.74	−2.27	4.45
SMR	8.20	9.58	7.26	9.78	9.27	11.63	7.09	18.54	5.03	14.69
TPSA	33	62	46	95	103	107	95	289	26	171
Density	0.73	0.85	0.67	0.75	0.72	0.76	0.74	0.94	0.66	0.89
vdw_area	269	326	299	395	389	454	259	723	206	554
vdw_vol	371	466	368	500	507	604	325	918	253	738
a_acc	2	3	2	4	2	3	2	11	1	6
a_don	0	1	1	2	0	0	1	6	0	3
a_hyd	13	18	12	17	16	22	8	31	7	26
KierA1	10.98	15.03	13.33	19.20	18.71	21.21	12.78	36.33	9.06	27.16
KierA2	4.63	6.50	6.15	9.61	9.19	10.70	5.38	18.18	3.63	12.46
KierA3	2.23	3.12	4.14	7.43	4.80	6.32	2.20	10.96	1.78	7.79
KierFlex	2.61	4.05	4.22	7.28	6.19	7.01	3.11	14.03	2.21	9.12

a) All drug clinically used in Japan.

Table 3. Enrichment Factors Attained for Four Drug Categories by the Drug-Like Profile Analysis

	Anxiolytic agent	β-Adrenergic blocker	ACE Inhibitor	Antibiotics
The number of active compounds in the training dataset	22	31	12	129
The number of active compounds in the test dataset	289	89	27	627
FP[a]	123	151	55	465
FP(active)[b]	46	72	14	271
EF(FP)	**10.6**	**44.1**	**77.6**	**7.6**
TR[c]	198	211	215	201
TR(active)[d]	47	53	20	115
EF(TR)	**6.8**	**23.2**	**28.3**	**7.5**
FP∪TR	287	290	238	560
FP∪TR(active)	67	74	20	289
EF(FP∪TR)	**6.6**	**23.6**	**25.6**	**6.8**
FP∩TR	34	72	32	106
FP∩TR(active)	26	51	14	97
EF(FP∩TR)	**21.8**	**65.5**	**133.3**	**12.0**

The enrichment factors (EFs) are shown in bold letters. *a*) The number of compounds predicted as active by molecular fingerprint. *b*) The number of known pharmacologically active compounds. *c*) The number of compounds predicted as active by trait. *d*) The number of known pharmacologically active compounds.

expense of low enrichment rate (EF(FP∪TR)) of 6.6. In this case almost 23% of the active compounds in the test dataset were covered. It is assumed that there are several reasons for the discrepancy between L_A and FP∩TR. The compounds in the training dataset are drugs clinically used now. The compounds in the test dataset, however, are compounds that show the pertinent activity but they are not used due to some reasons. This should be a main reason of the discrepancy. That is to say several highly specific molecules are used clinically.

In the case of β-adrenergic blocker, the number of active compounds in the test dataset is 89. The number of the top 2% of compounds with the highest drug probability is 211 of which 53 compounds are known as β-adrenergic blocker. The enrichment rate attained by use of the trait is 23.2. On the other hand, the number of compounds screened by use of molecular fingerprints was 151. The screened compounds contained 72 active compounds. The enrichment rate marked 44.1 in this case. If we get the union of the results by the two filters, 290 compounds were obtained and of which 74 compounds were active. The percentage of hit compounds is 26% and the enrichment rate is 23.6. The hit compounds almost cover 83% of the active compounds in the test dataset. The product of FP and TR gave 72 compounds of which 51 are experimentally verified active compounds. The enrichment factor reached 65.5 and about 57% of active compounds in the test dataset were covered. In this case, the current method showed a good performance.

The results for other two categories of compounds show that the enrichment rates attained by the current method are slightly variable depending on the drug category and the number of compounds in the dataset. In the case of ACE-inhibitor, a high enrichment factor of 133.3 was attained by use of FP∩TR. It covered 52% of the ACE-inhibitors in the test dataset. The enrichment factors for antibiotics, however, were relatively low. Since many different types of compounds that are active against various bacteria were contained in this category, the lower enrichment rate is not surprising.

Conclusions

Virtual screening is a method that complements current advances in high-throughput screening (HTS). It is highly desirable to screen as many compounds as possible in the early stage of drug discovery projects. Even with the modern technology of HTS, the number of compounds that can be handled is limited by several reasons. Virtual screening is undoubtedly becoming a computational counterpart of HTS now. In principle, there is no limit for the number of compounds handled by virtual screening. From the practical point of view the number of the compounds that can be managed by ordinary PC in laboratories could reach at least 10^3 times of the compounds that can be handled by HTS. Although virtual screening is becoming a practical tool in actual drug discovery projects, we have several serious problems to be solved. Using suitable criteria and filters, repeated searches are made through the virtual library with the goal to stepwise decrease its size. Since there is no unique strategy that leads the virtual screening process, there are a number of criteria that have to be considered in virtual screening. Virtual screening is applied mostly in a pragmatic way with the goal to speed up drug discovery in the most economic way. Excluding trivial compounds from the early stage of screening is one of the great benefits of virtual screening. The HTS or ordinary biological screening on especially focused sets of compounds selected by virtual screening can make the entire process of searching and optimizing leads cost-effective and fast. Since virtual screening is an entirely computer-based method, we can employ any sophisticated algorithms. In this paper, however, we used very simple and rather straightforward methods using the molecular fingerprints and trait that are observed in known active compounds. The reliability and enrichment rate could depend on the number of compounds in the training dataset. Therefore if we use a reliable proprietary dataset to train molecular fingerprint and trait that can characterize hit-likeness of compounds, the enrichment rate could increase. The results obtained in this paper, however, unequivocally have shown that the present method could be feasible in the actual drug discovery projects.

Acknowledgements One of the authors (N.H.) is grateful to the Research and Study Program of Tokai University Educational System General Research Organization for financial support.

References

1) "The Merck Index," 13th ed, Merck & Co. Inc., Whitehouse Station, U.S.A., 2001.
2) "Comprehensive Medicinal Chemistry," Elsevier MDL, San Ramon, U.S.A., 2003; http://www.mdli.com/
3) Pharmacast data, Technomics Inc., Tokyo, Japan, 2005.

4) McGregor M. J., Palli P. V., *J. Chem. Inf. Comput. Sci.*, **37**, 443—448 (1997).
5) Labute P., *J. Mol. Graphics. Mod.*, **18**, 464—477 (2000).
6) MOE: Molecular Operating Environment; Chemical Computing Group, Montreal, Canada; http://www.chemcomp.com/
7) Willett P., Winterman V. A., *Quat.Struc.-Act.Relat.*, **5**, 18—25 (1986).
8) Brooijmans N., Kuntz I. D., *Ann. Rev. Biophys. Biomol. Struct.*, **32**, 335—373 (2003).
9) Ryoka Systems Inc., Tokyo, Japan; http://www.rsi.co.jp/kagaku/cs/ccg/dl/index.html

Megsin Gene: Its Genomic Analysis, Pathobiological Functions, and Therapeutic Perspectives

Toshio Miyata[1,*], Ming Li[2], Xueqing Yu[2] and Noriaki Hirayama[3]

[1]*Division of Nephrology, Hypertension and Metabolism, and* [3]*Basic Medical Science and Molecular Medicine Tokai University School of Medicine, Kanagawa, Japan, and* [2]*Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, P.R. China*

Abstract: It is critical to uncover genes specifically expressed in individual cell types for further understanding of cell biology and pathology. In order to elucidate pathogenesis of renal disease, we performed functional quantitative analysis of the genome in human kidney cells and compared the expression levels of a variety of kidney transcripts with those in other non-kidney cells. As a result, we identified a novel human gene, megsin, which is a new serine protease inhibitor (serpin) predominantly expressed in the kidney. Megsin is up-regulated in kidney disease. Genomic analysis revealed an association of the polymorphisms of megsin gene with susceptibility and/or progression of kidney disease. Its overexpression in rodents has led to the recognition of two different kidney abnormalities. The first disorder is linked to megsin biological effect itself and the other to its conformational abnormality recently called the serpinopathy. In the latter model, the cellular and tissue damage is induced by the endoplasmic reticulum (ER) stress due to conformational disorder resulting from megsin tertiary structure. In both types, the inhibition of megsin's activity or abnormal conformational change should open new therapeutic perspectives. The desire to prevent these abnormalities with the hope to offer new therapeutic strategies has stimulated the development of megsin inhibitors by a structure based drug design approach relying on a precisely known three dimensional megsin structure.

Received on: February 12, 2007 - Revised on: February 13, 2007 - Accepted on: March 27, 2007

Key Words: Tissue specific gene, renal disease, serpin, serpinopathy, endoplasmic reticulum stress, structure based drug design (SBDD).

ISOLATION OF A KIDNEY-SPECIFIC GENE, MEGSIN

Sixty trillion (6×10^{13}) cells in the human body share essentially identical genomic DNA. Nonetheless, for normal physiological function distinct for each cell type, expression of genes is tightly regulated by the cell lineage. It is therefore critical to uncover which genes are specifically expressed in individual cell types for further understanding of cell biology and pathology of diseases.

Some molecular biological approaches such as subtraction cloning and differential hybridization allowed us to compare libraries from two different sources and detect cell- or organ-specific genes. However, no quantitative information about expression levels of specific and non-specific genes can be obtained. No more than two libraries can be compared at one time using these methods, either.

In order to elucidate pathogenesis of kidney disease, we performed functional quantitative analysis of the genome in human kidney cells and compared the expression levels of a variety of kidney transcripts with those in other non-kidney cells. For this purpose, we employed a unique strategy utilizing a rapid large scale sequencing of a 3'-directed cDNA library and computerized data processing [1]. As kidney cells, we focused on mesangial cells, as they play a central role in maintaining a structure and function of the kidney glomerulus. In the pathophysiology of kidney disease, the proliferation of mesangial cells and the accumulation of extracellular mesangial matrix are primary events leading to the progression of a variety of kidney diseases such as chronic glomerulonephritis and diabetic nephropathy, two major causes of end-stage renal failure.

Sequencing of a 3'-directed cDNA library of cultured human mesangial cells was performed to determine partial sequences of 1836 randomly picked clones [2]. The sequence similarities of the clones were compared with each other and with the DNA databank GenBank utilizing the FASTA program [3]. We compared their transcripts with those in various other cells and organs [2], and finally selected six clones [4, 5], which had not been recorded in any other databases and were thought to be predominantly expressed in mesangial cells. Dot blot anlaysis of mRNAs from various organs and cells supported mesangium-predominant expression of these clones. One clone was obtained which was most abundant among those detected only in our human mesangial cell cDNA library (0.3 % of all the mRNA population). Utilizing PCR techniques, a full length cDNA of this gene was obtained [4].

MEGSIN GENE AND PROTEIN

Amino acid homology search by the FASTA program throughout the SwissProt database revealed that this gene

*Address correspondence to this author at the Division of Nephrology, Hypertension and Metabolism, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan; Tel: +81 463 93 1936; Fax: +81 463 93 1938; E-mail: t-miyata@is.icc.u-tokai.ac.jp

1389-2029/07 $50.00+.00

was highly homologous to members of the serpin superfamily [6], which is a group of structurally related proteins and generally serves as extracellular, irreversible serine protease inhibitors. We therefore termed this gene 'megsin' after mesangial cell-predominant gene with a homology to serpin.

The amino acid sequence in the reactive loop site of megsin exhibits the characteristic features of functional serpins [4]. Our *in vitro* assays utilizing recombinant megsin indeed confirmed that megsin serves as a functional serpin [7].

EXPRESSIONS OF MEGSIN GENE AND PROTEIN IN THE KIDNEY

Northern blot and reverse-transcribed polymerase chain reaction analyses of various tissues and cells demonstrated that megsin was predominantly expressed in human mesangial cells [4]. These findings were further confirmed by *in situ* hybridization and by immunohistochemistry (Fig. **1**) using megsin-specific antibodies [4, 8, 9]. In IgA nephropathy and diabetic nephropathy, megsin mRNA expression in glomeruli was up-regulated. A similar up-regulation of megsin was observed in the experimental anti-Thy1 nephritis model of rats [10]. The increased expression of megsin gene is thus associated with renal disorders with mesangial proliferation and its matric accumulation.

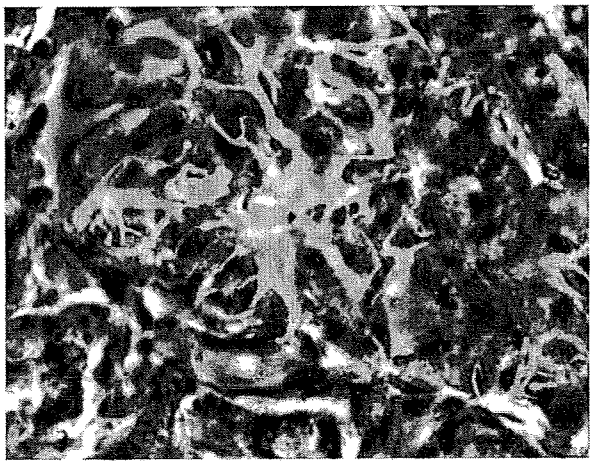

Fig. (1). Megsin protein expression in the kidney glomerulus. Immunofluorescence study utilizing anti-human megsin demonstrates that megsin is predominantly localized in the glomerulus, especially in the mesaigial area (x 200).

GENOMIC ASSOCIATION OF MEGSIN WITH KIDNEY DISEASE

Recent studies have demonstrated the interesting association of the polymorphisms of megsin gene with susceptibility and/or progression of kidney disease in Chinese patients [11-13].

The correlation between polymorphisms of megsin gene and IgA nephropathy were investigated by using the family-based association study. Polymorphisms C2093T and C2180T within the 3' untranslated region of megsin were first examined. Transmission disequenlibrium test (TDT) and haplotype relative risk (HRR) analyses revealed that megsin 2093C and 2180T alleles were significantly more transmitted from heterozygous parents to patients, which suggested that the genetic variation in megsin conferred susceptibility to IgA nephropathy [11].

To further examine the relationships of these genetic variants with clinical manifestations and renal histological lesions, haplotypes were constructed by using the C2093T and C2180T alleles. The genotype-phenotype relationship study found that the 2093C-2180T haplotype is associated with more severe forms of IgA nephropathy and more rapid disease progression [12]. It raised the question that whether these two variants confer the effect or just in linkage disequilibrium with other variants nearby. To answer this question, 12 known SNPs from different functional regions of megsin were selected from GenBank. The genotypes were determined by PCR-RFLP and direct sequencing and the heterozygosis rates were calculated if the genotypes were heterozygote. When the rate exceeded 10 %, the TDT and HRR analysis were performed. We found two novel SNPs which hadn't been reported before (23179 9T/10T and 23103 7A/6A), and six heterozygous SNPs, among which five SNPs with the rate more than 10 % were analyzed. TDT and HRR analyses showed that 23167G alleles were transmitted more frequently from parents to patients than expected. The scores of glomerular index and glomerular sclerosis index were higher in GG genotype patients than those in other genotypes and the distribution frequency of GG genotype in the progressive group was higher than that of the stable group. The polymorphism of megsin A23167G is thus associated with susceptibility and progression of IgA nephropathy in Chinese population. GG genotype is associated with severe histological lesions and progression of the disease [13]. The analysis of other four SNPs found no statistical significance. These data suggest the possible involvement of genetic variations of megsin in the susceptibility and progression of IgA nephropathy.

PATHOBIOLOGICAL FUNCTION OF MEGSIN

To further understand a pathobiological role of megsin, we overexpressed the human megsin cDNA in mouse genome [7]. Two lines of megsin transgenic mice have been obtained. They developed progressive mesangial matrix accumulation, an increase in the number of mesangial cells (proliferation), and an augmented immune complex deposition (Figs. **2A** and **B**). The transgenic model is characterized by the expression of megsin in all tissues due to the ubiquitous promoter for the transgene. Although immunohistochemical studies revealed the presence of megsin in a host of tissues as well as in non-mesangial areas of the kidney, pathogenic effects of megsin overexpression were restricted within glomeruli. The mechanism of glomerular abnormalities still remains unknown. We speculate that overexpression of megsin, a novel serpin expressed in the glomerulus, may lead to mesangial dysfunction, impair the disposal of immune complexes, and increase mesangial matrix by tipping the balance towards lower matrix degradation. By contrast, histological abnormalities were not evident in knock-out mice for megsin (our unpublished observation).

We subsequently cross-bred megsin transgenic mice with RAGE (the receptor for advanced glycation end products)/iNOS double transgenic diabetic mice [14], and produced a severe diabetic nephropathy model [15]. These triple transgenic mice were the first to exhibit diabetic nodular lesions, the end result of mesangial matrix accumulation, similar to those observed in humans (Fig. **2C**). A genetic manipulation of megsin thus engenders kidney diseases, suggesting its biopathological involvements.

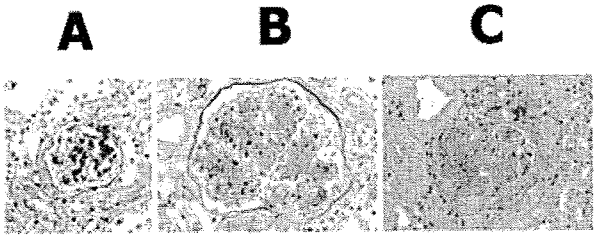

Fig. (2). Histopathological analysis of megsin transgenic mouse kidneys (PAS staining). As compared to 40-week-old wild-type mice (A, x 200), F_1 megsin transgenic mouse of the same age developed mesangial matrix expansion and an increase in the number of mesangial cells (B, x 200). This megsin transgenic mice were cross-bred with RAGE/iNOS transgenic mice to generate the triple transgenic mice, which developed more severe glomerular lesion such as glomerular hypertrophy, global mesangial expansion, adhesion of parietal epithelial cells to the tuft, nodular-like lesions (C, x 200).

SERPINOPATHY: A NOVEL PATHOLOGY INDUCED BY MEGSIN GENE OVEREXPRESSION

As an extension of our search, we generated megsin transgenic rats utilizing the same ubiquitous promoter for the transgene [16]. Homozygotes had an impaired growth. They failed to gain weight and all died within 10 weeks. Western blot analysis revealed that megsin expression is ubiquitous in all tissues with different degrees: high in heart, kidney and pancreas, and moderate in lung. Homozygotes had clear evidence of renal and pancreatic dysfunction, as shown by a nephrotic syndrome (proteinuria, hypoproteinemia, and elevated cholesterol levels) with deterioration of renal function (creatinine and BUN) and hyperglycemia with insulin deficiency. Compared to size-matched wild-type rats, they have a markedly lower creatinine clearance, a marker of renal function.

By histology, numerous large Periodic Acid Schiff (PAS)-positive droplets were observed within the cytoplasm of kidney and pancreatic cells (Fig. **3**). In heterozygotes, a significantly lower number of smaller PAS-positive droplets were observed with essentially the same distribution pattern.

Many questions beg for an answer. Why is kidney pathology so different between mice and rats overexpressing megsin while we utilize the same promotor? Why is diabetes associated only with megsin transgenic rats while it is absent in mice? What is the molecular identity of the PAS-positive droplets and electron-dense inclusion bodies within the ER?

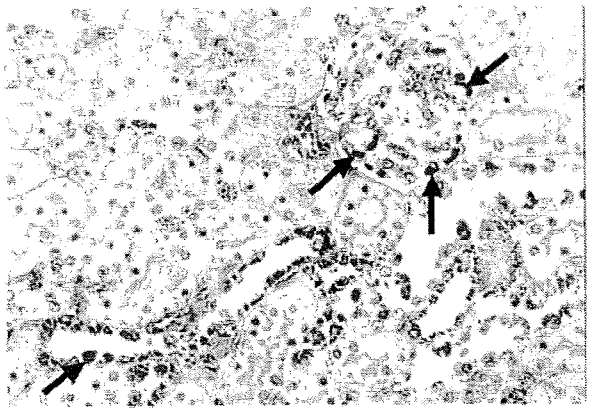

Fig. (3). PAS staining of a megsin transgenic rat (homozygote) at the age of 8 weeks. Numerous, large, PAS-positive droplets were observed within the cytoplasm of glomerular epithelial cells, distal tubular cells, and collecting ducts. x 200.

Fortunately, the clinical literature provided us a stimulating model [17]. Patients deficient in $alfa_1$-antitrypsin (AT), another serpin, suffer from pulmonary emphysema. Some suffer from additional liver cirrhosis. In the latter patients, the mutated $alfa_1AT$ molecule is retained within the liver endoplasmic reticulum (ER) as PAS-positive but diastase-resistance droplets, meaning that they are not glycogen [18]. Recent advances in molecular analysis unraveled this pathology. A mutation of $alfa_1AT$ enlarges the strand 4 pocket, thereby facilitating an aberrant intermolecular linkage between the loop of one $alfa_1AT$ molecule and the strand 4 pocket of another $alfa_1AT$ molecule. The resulting dimmer eventually polymerizes within the ER. The polymers cannot be exported into other organellas such as Golgi, accumulate within cells as PAS-droplets, and trigger a series of hepatotoxic events [19].

The pathology underlying pulmonary emphysema and liver cirrhosis in $alfa_1AT$ deficiency is thus quite different. A loss-of-function of $alfa_1AT$ reduces the inactivation of elastase, the target protease, with attendant pulmonary emphysema. By contrast, a gain-of-function due to the aggregation of mutated $alfa_1AT$ within the ER induces liver dysfunction. The former is a biological disorder and the latter a conformational disorder.

This pathology has been recently referred to as the serpinopathy [17]. Two serpinopathies have been documented to date in humans: liver cirrhosis due to the mutation of $alfa_1AT$ and familial encephalopathy due to neuroserpin [20]. Conformational alterations of protein tertiary structure involving beta sheet are already well known as they are at the root of prions, amyloidosis and serpinopathy [21].

The disorders observed in our megsin transgenic rats are caused by megsin polymerization and may be called a serpinopathy. On electron microscopy, indeed the huge electron-dense inclusions positive for megsin accumulate in the dilated rough ER [16]. Such a histological abnormality of ER is absent in megsin transgenic mice. Our model of megsin serpinopathy is the first to involve the kidney and pancreas.

The phenotypic difference between megsin transgenic mice and rats might reflect different pathogenic mechanisms depending upon the degrees of megsin gene expression. In mice, overexpression of megsin only enhances the inactivation of its target serine protease. Histological abnormalities develop late at the age of 40 weeks. Megsin polymerization is absent in this model. In rats, by contrast, the marked overexpression of megsin is one order magnitude higher than in mice. It induces early megsin polymerization and cellular toxicity within 10 weeks.

ENDOPLASMIC RETICULUM (ER) STRESS

Why is megsin overexpression toxic only for the kidney and pancreas while megsin is also overexpressed in other organs? What are the mechanism(s) underlying the cytotoxicity and tissue damage associated with the serpinopathy?

We took advantage of our model to elucidate the mechanism of cellular toxicity in serpinopathies. We demonstrated that an ER stress induced by megsin polymerization initiates cellular toxicity. ER is an intracellular compartment that plays a critical role in the processing, folding and transport of newly synthesized proteins [22]. All cells regulate the capacity of their ER to process synthesized proteins and adapt to an imbalance between protein load and folding capacity, recently referred to as the ER stress [23, 24]. ER stress is triggered by various stimuli and pathophysiological conditions [25, 26]. As a defensive system against the ER stress, an unfolded protein response (UPR) develops [27]. UPR involves transient attenuation of new protein synthesis, degradation of misfolded proteins, expression of a variety of ER stress proteins such as oxygen-regulated protein (ORP) 150 [28] and glucose-regulated proteins (GRPs). Under normal conditions, these ER stress inducible proteins serve as protein chaperones, complex with defective proteins and target them for degradation. During stress, UPR may limit accumulation of abnormal proteins within the ER, allowing cells to tolerate the ER stress. When the ER stress exceeds the balance beyond the limit of cellular UPR, the cell undergoes the apoptosis by activating caspase 12 and CHOP. ER is thus a very interesting organella because it is a crucial center for the maintenance of life by protein synthesis as well as for cell death [29].

On immunohistochemistry, ER stress inducible chaperons, ORP150 and GRP 78, were markedly up-regulated in glomeruli of megsin transgenic rats as compared to wild-type [30]. By double immunostaining, OPR150-expressed podocytes showed accumu-lation of megsin inclusion bodies. Increased expression of ORP150 was confirmed by Western blot analysis. Up-regulation was markedly in homozygotes, mildly in heterozygotes, but absent in wild-type kidney. Interestingly, megsin expression was equally abundant in the heart in the absence of PAS droplets and electron dense-inclusions. Indeed, ER stress inducible chaperone were not detectable in the heart [30].

Altogether, in transgenic rats, the aggregation of megsin as a consequence of aberrant intermolecular linkage within the ER perturbs the function of the ER and an unfolded protein response ensues. Despite an increased expression of ER stress inducible chaperons, some cells are damaged and undergo apoptosis.

MECHANISM OF MEGSIN'S ACTION AND ITS THERAPEUTIC PERSPECTIVE

The analysis of genes expressed in kidney mesangial cells has thus allowed the identification of a new kidney specific serpin, megsin. Its overexpression in rodents has led to the recognition of two different kidney abnormalities. The first disorder is linked to megsin biological effect itself and the other to its conformational abnormality recently called the serpinopathy. In both types, the inhibition of megsin's activity or abnormal conformational change should open new therapeutic perspectives. As yet, however, only very few serpin inhibitors have been reported and none is in clinical use.

In the last part of this article, I discuss strategies designed to inhibit megsin's activity or abnormal conformational change leading to the serpinapathy.

The mechanism of megsin's activity is fortunately deduced and might provide a clue to develop such an inhibitor. All serpins share a common structure [31]: a beta sheet-rich body and an exposed mobile reactive loop which functions as a pseudo-substrate for the target protease (Fig. **4**). The target protease first cleaves the loop which traps the protease. The cleaved loop then penetrates into a pocket, named the strand 4 position in the A beta sheets. This step is crucial for the serpin's anti-protease activity, because, without insertion of the cleaved loop into this pocket, the trapped protease is released from the loop. Indeed, pre-incubation of a megsin with a 14 amino acid peptide of the loop abolishes the megsin's activity [our unpublished observation].

Identification of low molecule compounds able to enter into this pocket as a mock molecule may therefore be a key to develop a serpin inhibitor. Of note, the loop amino acid content and hence the pocket size and charge differs in each serpin, potentially determining its specificity.

We also have a clue to prevent the progression of the megsin serpinopathy, because identification of such a low molecule compound able to enter into the strand 4 pocket may be also a key to prevent aberrant intermolecular linkage between the loop and the pocket.

STRUCTURE BASED DRUG DESIGN FOR DEVELOPMENT OF MEGSIN INHIBITORS

Up to now, very few serpin inhibitors have been reported. Most have been discovered by high-throughput random screening of a large chemical library [32-34], a rather inefficient strategy. We tried a new approach, the structure based drug design (SBDD). This approach requires a precise knowledge of the three dimensional structure of the serpin. As the megsin crystal structure was not available, I chose another clinically important serpin, plasminogen activator inhibitor (PAI-1), whose crystal structure has been described [35], and tested our strategy with SBDD.

We virtually screened a library, encompassing more than two millions chemicals by SBDD (Fig. **5**). Two different filters reduced the number of compounds to about 3,000. The first filter assesses the drug-likeness, calculated from the specific distributions of the molecular descriptors for drug molecules clinically used. The second filter assesses the spe-

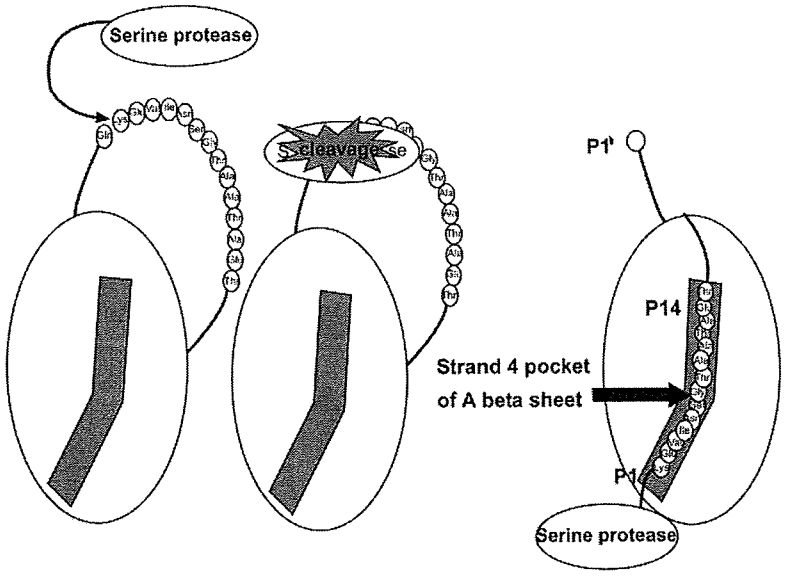

Fig. (4). Mechanism of serpin's action and conformational change.

Serpins consist of a beta sheet-rich body. It contains an exposed mobile reactive loop. The target protease first cleaves the loop. The cleaved loop then traps the protease and inserts its *N*-terminus part into the strand 4 position (s4A) of the A beta sheet, triggering its anti-protease activity. Occasionally, an aberrant intermolecular linkage induced by the insertion of serpin's loop into the s4A position of another serpin molecule results in intracellular aggregation and subsequent cellular damage, called the serpinopathy. Therefore, small loop peptide or compounds able to enter into the s4A position of the A beta sheet as a mock molecule may prevent the serpin biological activity or an aberrant intermolecular linkage of serpins leading to the serpinopathy.

cific lead-likeness of the inhibitory molecules, calculated from distributions of the molecular descriptors common to known reference inhibitors and the inhibitory 14 amino acid peptide to PAI-1. Docking simulation was then undertaken

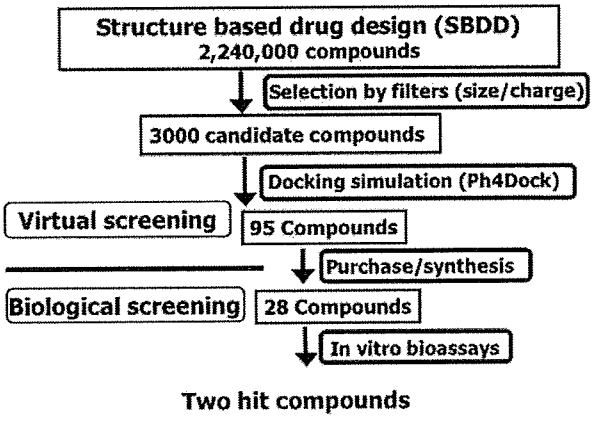

Fig. (5). Structure based drug design (SBDD) approach.

by a program Ph4Dock [36] to evaluate whether the compound fits within the PAI-1 pocket. Eventually, we identified 95 candidate compounds theoretically able to bind this pocket and purchased or synthesized 28 of them to test their biological activities *in vitro*. Two of the 28 candidate compounds were highly effective [our unpublished observation]. Their PAI-1 inhibitory activity appeared specific as they fail to modify other serpin/serine protease systems. On SDS-PAGE, PAI-1 formed a covalent complex with tPA whereas no PAI-1/tPA complex formation was observed when PAI-1 was preincubated with our PAI-1 inhibitory compounds. This, we have succeeded in isolation of a low molecular inhibitory compound for PAI-1, a model serpin.

Based upon these results, we are trying to elucidate the megsin crystal structure in order to analyze its strand 4 pocket to identify a compound potentially able to enter into this pocket. Such a compound will hopefully prevent the consequences of experimental kidney diseases as well as the serpinopathy phenotype of our megsin transgenic rat model.

SUMMARY

The analysis of genes expressed in human kidney cells has allowed the identification of a new kidney specific serpin, megsin. It is up-regulated in human kidney diseases. Furthermore, recent genomic analysis revealed an association of the polymorphisms of megsin gene with susceptibility and/or progression of kidney disease. Interestingly, its overexpression in rodents has led to the recognition of two different kidney abnormalities. One of which (megsin transgenic rat) has provided a major clue to understand a new family of clinical diseases, called the serpinopathy. The desire to prevent these abnormalities with the hope to offer new therapeutic strategies has stimulated the development of new megsin inhibitors by a structure based drug design approach relying on a precisely known dimensional megsin structure. Preliminary results obtained for PAI-1, another serpin, are very promising.

ACKNOWLEDGEMENTS

This study was supported by grants from the New Energy and Industrial Technology Development Organization and from the Program for Promotion of Fundamental Studies in Health Sciences of the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan to T.M.

REFERENCES

[1] Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., Matsubara, K. *Nature Genet.* **1992** , *2*: 173.
[2] Yasuda, Y., Miyata, T., Nangaku, M., Iida, Y., Maeda, K., Kurokawa, K., Okubo, K. *Kidney Int.* **1998** , *53*: 154.
[3] Pearson, W.R., Lipman, D.J. *Proc. Natl. Acad. Sci.* **1988**, *85*: 2444.
[4] Miyata, T., Nangaku, M., Suzuki, D., Inagi, R., Uragami, K., Sakai, H., Okubo, K., Kurokawa, K. *J. Clin. Invest.* **1998**, *102*: 828.
[5] Wada, T., Miyata, T., Inagi, R., Nangaku, M., Wagatsuma, M., Suzuki, D., Wadinski, B.E., Okubo, K., Kurokawa, K. *J. Am. Soc. Nephrol.* **2001** , *12*: 2601.
[6] Carrell, R.W., Boswell, D.R. *Elsevier Science.* **1986** , 403.
[7] Miyata, T., Inagi, R., Nangaku, M., Imasawa, T., Sato, M., Izuhara, Y., Suzuki, D., Yoshino, A., Onogi, H., Kimura, M., Sugiyama, S., Kurokawa, K. *J. Clin. Invest.* **2002** , *109*: 585.
[8] Suzuki, D., M, T., Nangaku, M., Takano, H., Saotome, N., Toyoda, M., Mori, Y., Zhang, S., Inagi, R., Endoh, M., Kurokawa, K., Sakai, H. *J. Am. Soc. Nephrol.* **1999**, *10*: 2606.
[9] Inagi, R., Miyata, T., Suzuki, D., Toyoda, M., Wada, T., Ueda, Y., Izuhara, Y., Sakai, H., Nangaku, M., Kurokawa, K. *Biochem. Biophys. Res. Commun.* **2001**, *286*: 1098.
[10] Nangaku, M., Miyat, T., Suzuki, D., Umezono, T., Hashimoto, T., Wada, T., Yagi, M., Nagano, N., Inagi, R., Kurokawa, K. *Kidney Int.* **2001**, *60*: 641.
[11] Li, Y.J., Du, Y., Li, C.X, Guo, H., Leung, J.C., Lam, M.F., Yang, N., Huang, F., Chen, Y., Fang, J.Q., Maxwell, P.H., Lai, K.N., Wang, Y. *J. Am. Soc. Nephrol.* **2004** , *15*: 1739.
[12] Xia, Y.F., Li, Y.J., Du, Y., Yang, N.S., Li, C.X., Leung, J.C., Lam, M.F., Huang, W.J., Chen, S.Q., Maxwell, P.H., Lai, K.N., Wang, Y.M. *Nephrol. Dial. Transplant.* **2006**, *21*: 1570
[13] Xia, Y.F., Huang, S., Li, X., Yang, N., Huang, J., Xue, C., Zhang, M., Leung, J.C., Lam, M.F., Li, J. *Clin. Nephrol.* **2006**, *65*: 153.
[14] Yamamoto, Y., Kato, I., Doi, T., Yonekura, H., Ohashi, S., Takeuchi, M., Watanabe, T., Yamagishi, S., Sakurai, S., Takasawa, S., Okamoto, H., Yamamoto, H. *J. Clin. Invest.* **2001**, *108*: 261.
[15] Inagi, R., Yamamoto, Y., Nangaku, M., Usuda, N., Okamoto, H., Kurokawa, K., van, Ypersele, de, Strihou, C., Yamamoto, H., Miyata, T. *Diabetes* **2006**, *55*: 356.
[16] Inagi, R., Nangaku, M., Usuda, N., Shimizu, A., Onogi, H., Izuhara, Y., Nakazato, K., Ueda, Y., Oishi, H., Takahashi, S., Yamamoto, M., Suzuki, D., Kurokawa, K., van, Ypersele, de, Strihou, C., Miyata, T. *J. Am. Soc. Nephrol.* **2005**, *16*: 1339.
[17] Lomas, D.A., Carrell, R.W. *Nat. Genet. Rev.* **2002**, *3*: 759.
[18] Lomas, D. A., Evans, D.L., Finch, J.T., Carrell, R. W. *Nature* **1992**, *357*: 605.
[19] Lomas, D.A., Mahadeva, R. *J. Clin. Invest.* **2002**, *110*: 1585.
[20] Davis, R.L., Shrimpton, A.E., Holohan, P.D., Bradshaw, C., Feiglin, D., Collins, G.H., Sonderegger, P., Kinter, J., Becker, L.M., Lacbawan, F., Krasnewich, D., Muenke, M., Lawewnce, D.A., Yerby, M.S., Shaw, C.M., Gooptu, B., Elliott, P.R., Finch, J.T., Carrell, R.W., Lomas, D.A. *Nature* **1999**, *401*: 376.
[21] Carrell, R.W., Lomas, D.A. *Lancet* **1997**, *350*: 134.
[22] Ellgaard, L., Helenius, A. *Curr. Opin. Cell. Biol.* **2001**, *13*: 431.
[23] Lee, A.S. *Trends Biochem Sci.* **2001**, *26*: 504.
[24] Lee, A.S. *Curr. Opin. Cell Biol.* **1992**, *4*: 267.
[25] Aridor, M., Balch, W.E. *Nat. Med.* **1999**, *5*: 745.
[26] Ron, D. *J. Clin. Invest.* **2002**, *109*: 443.
[27] Kaufmann, R.J. *Genes Dev.* **1999**, *13*: 1211.
[28] Kitao, Y., Ozawa, K., Miyazaki, M., Tamatani, M., Kobayashi, T., Yanagi, H., Okabe, M., Ikawa, M., Yamashima, T., Stern, D.M.; Hori, O., Ogawa, S. *J. Clin. Inves.* **2001**, *108*: 1439.
[29] Pahl, H.L. *Physiol. Rev.* **1999**, *79*: 683.
[30] Inagi, R., Nangaku, M., Onogi, H., Ueyama, H., Kitao, Y., Nakazato, K., Ogawa , S., Kurokawa, K., Couser, W.G., Miyata, T. *Kidney Int.* **2005**, *68*: 2639.
[31] Gils, A., Declerck, P. *J. Thromb. Haemost.* **2004**, *91*: 425.
[32] Elokdah, H., Abou-Gharbia, M., Hennan, J.K., McFarlane, G., Mugford, C.P., Krishnamurthy, G., Crandall, D.L. *J. Med. Chem.* **2004**, *47*: 3491.
[33] Crandall, D.L., Elokdah, H., Di, L., Hennan, J.K., Gorlatova, N.V., Lawrence, D.A. *J. Thromb. Haemost.* **2004**, *2*: 1422.
[34] Liang, A., Wu, F., Tran, K., Jones, S.W., Deng, G., Ye, B., Zhao, Z., Snider, R.M., Dole, W.P., Morser, J., Wu, O. *Thromb. Res.* **2005**, *115*: 341.
[35] Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. *J. Mol. Biol.* **1997**, *112*: 535.
[36] Goto, J., Kataoka, R., Hirayama, N. *J. Med. Chem.* **2004**, *47*: 6804.

A Novel Class of Prolyl Hydroxylase Inhibitors Induces Angiogenesis and Exerts Organ Protection Against Ischemia

Masaomi Nangaku, Yuko Izuhara, Shunya Takizawa, Toshiharu Yamashita, Yoshiaki Fujii-Kuriyama, Osamu Ohneda, Masayuki Yamamoto, Charles van Ypersele de Strihou, Noriaki Hirayama, Toshio Miyata

Objective—Hypoxia inducible factor (HIF) plays a pivotal role in the adaptation to ischemic conditions. Its activity is modulated by an oxygen-dependent hydroxylation of proline residues by prolyl hydroxylases (PHD).

Methods and Results—We discovered 2 unique compounds (TM6008 and TM6089), which inhibited PHD and stabilized HIF activity in vitro. Our docking simulation studies based on the 3-dimensional structure of human PHD2 disclosed that they preferentially bind to the active site of PHD. Whereas PHD inhibitors previously reported inhibit PHD activity via iron chelation, TM6089 does not share an iron chelating motif and is devoid of iron chelating activity. In vitro Matrigel assays and in vivo sponge assays demonstrated enhancement of angiogenesis by local administration of TM6008 and TM6089. Their oral administration stimulated HIF activity in various organs of transgenic rats expressing a hypoxia-responsive reporter vector. No acute toxicity was observed up to 2 weeks after a single oral dose of 2000 mg/kg for TM6008. Oral administration of TM6008 protected neurons in a model of cerebrovascular disease. The protection was associated with amelioration of apoptosis but independent of enhanced angiogenesis.

Conclusions—The present study uncovered beneficial effects of novel PHD inhibitors preferentially binding to the active site of PHD. (*Arterioscler Thromb Vasc Biol.* **2007;27:2548-2554.**)

Key Words: hypoxia ■ hypoxia inducible factor ■ structure based drug design ■ stroke ■ ischemia

Oxygen supply declines under ischemic conditions in many human vascular diseases including ischemic heart disease, chronic kidney failure, and stroke. The resulting hypoxia causes functional impairment of cells as well as structural tissue damage and triggers a broad spectrum of cellular defenses such as angiogenesis, erythropoiesis, glycolysis, and antioxidative enzymes.

Hypoxia-inducible factor (HIF), a heterodimeric nuclear factor, is a crucial intermediate in these defensive mechanisms.[1–3] Under normoxic conditions, HIF is constitutively transcribed and translated. Its stability is drastically reduced by the oxygen-dependent enzymatic hydroxylation of proline residues by prolyl hydroxylases (PHD).[4–9] Hydroxylated HIF recruits the E3-ubiquitin ligase, von Hippel Lindau protein (pVHL)[10,11] which, in turn, tags HIF with ubiquitin groups and targets it for degradation by the proteasome.[12,13] Under hypoxic conditions, HIF is not hydroxylated but binds to its heterodimeric partner HIF-1β. The resulting protein complex transactivates in the nucleus a host of genes involved in the adaptation to hypoxic stress.[14]

Activation of HIF may prove therapeutic for vascular disorders. Most treatments for ischemic and hypoxic disorders are currently focused on symptomatic relief and correction of etiologic factors. Drugs dissolving thrombi are also used to restore blood flow in the acute phase. As yet no compound enhancing organ resistance to hypoxia is clinically available. HIF activates a "master gene" switch that results in a broad and coordinated downstream reaction, protecting tissues against the consequences of hypoxia. The availability of less cumbersome non-toxic small molecular activators of HIF should prove very useful for therapeutic intervention.[15,16]

To obtain such novel compounds and to understand a molecular mechanism of PHD inhibition, we performed docking simulation based on the 3-dimensional structure of human PHD2. We further documented the in vitro and in vivo effectiveness of the novel PHD inhibitors we identified.

Materials and Methods

Please see the supplemental data section at http://atvb.ahajournals.org for detailed Methods.

Original received May 24, 2007; final version accepted September 26, 2007.

From the Division of Nephrology and Endocrinology (M.N.), University of Tokyo School of Medicine, Japan; the Institute of Medical Sciences (Y.I., S.T., T.M.), Divisions of Nephrology, Hypertension, and Metabolism and of Neurology, Tokai University School of Medicine, Kanagawa, Japan; the Center for Tsukuba Advanced Research Alliance and Institute of Basic Medical Sciences (T.Y., O.O.), University of Tsukuba, Japan; the Center for Tsukuba Advanced Research Alliance (Y.F.-K.), University of Tsukuba, Japan; the Center for Tsukuba Advanced Research Alliance and JST-ERATO Environmental Response Project (M.Y.), University of Tsukuba, and the Department of Medical Biochemistry, Tohoku University Graduate School of Medicine, Sendai, Japan; the Service de Nephrologie (C.v.Y.d.S.), Universite Catholique de Louvain, Brussels, Belgium; and the Basic Medical Science and Molecular Medicine (N.H.), Tokai University School of Medicine, Kanagawa, Japan.

Correspondence to Toshio Miyata, MD, PhD, Institute of Medical Sciences and Division of Nephrology, Hypertension and Metabolism, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan. E-mail t-miyata@is.icc.u-tokai.ac.jp

Arterioscler Thromb Vasc Biol is available at http://atvb.ahajournals.org

DOI: 10.1161/ATVBAHA.107.148551