PROTEINS: Structure, Function, and Genetics 23:536–547 (1995)

Atomic and Residue Hydrophilicity in the Context of Folded Protein Structures

Leslie A. Kuhn,[1] Craig A. Swanson,[2] Michael E. Pique,[2] John A. Tainer,[2] and Elizabeth D. Getzoff[2]
[1]Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824 and [2]Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037.

ABSTRACT **Water–protein interactions drive protein folding, stabilize the folded structure, and influence molecular recognition and catalysis. We analyzed the closest protein contacts of 10,837 water molecules in crystallographic structures to define a specific hydrophilicity scale reflecting specific rather than bulk solvent interactions. The tendencies of different atom and residue types to be the nearest protein neighbors of bound water molecules correlated with other hydrophobicity scales, verified the relevance of crystallographically determined water positions, and provided a direct experimental measure of water affinity in the context of the folded protein. This specific hydrophilicity was highly correlated with hydrogen-bonding capacity, and correlated better with experimental than computationally derived measures of partitioning between aqueous and organic phases. Atoms with related chemistry clustered with respect to the number of bound water molecules. Neutral and negatively charged oxygen atoms were the most hydrophilic, followed by positively-charged then neutral nitrogen atoms, followed by carbon and sulfur atoms. Agreement between observed side-chain specific hydrophilicity values and values derived from the atomic hydrophilicity scale showed that hydrophilicity values can be synthesized for different functional groups, such as unusual side or main chains, discontinuous epitopes, and drug molecules. Two methods of atomic hydrophilicity analysis provided a measure of complementarity in the interfaces of trypsin:pancreatic trypsin inhibitor and HIV protease:U-75875 inhibitor complexes.**

Key words: water, hydrophobicity, hydration, X-ray crystallography, solvation, ordered solvent, molecular recognition, water–protein interactions, drug and inhibitor design, protein surface analysis

INTRODUCTION

Interactions between amino acid residues and their aqueous and protein environments together determine protein folding and mediate intermolecular interactions.[1–7] Protein–water stereochemistry and hydrogen bonding have been extensively analyzed in crystallographic and NMR structures[1,8–15] and recently reviewed.[16] Because protein crystals contain 27–77% water,[17,18] surface-exposed residues (not in protein–protein interfaces) in these crystals are well-solvated. While bulk water surrounds all surface-exposed atoms in proteins, bound water molecules, whose time-averaged positions are measured by crystallography, represent persistent, energetically favored interactions between solvent and protein.[19] Here, we develop "specific hydrophilicity" scales reflecting specific water interactions, rather than bulk solvation, from the positions of water molecules bound to protein surfaces in 56 X-ray crystallographic structures. We compare these new scales with hydrophilicity and hydrophobicity scales based on different properties of amino acids and their analogs, such as solvent-exposed surface area,[20–22] parametric fits to free energy or surface exposure data,[23–26] and preference for aqueous versus organic phases.[27–29] The high correlation observed between the side-chain and residue crystallographic specific hydrophilicity scales and other experimentally determined hydrophilicity scales dispels the commonly held view that crystallographic surface-bound water is primarily determined by the crystal lattice rather than specific chemical interactions.

The atomic specific hydrophilicity values for every atom type occurring in proteins (e.g., Ala C_α, Ala C_β, . . .) were determined to identify atom types with similar hydration and to provide an atomic hydrophilicity scale appropriate for detailed analysis of protein interactions. We then evaluated the influence of intraprotein hydrogen bonds and salt bridges on the relative hydrophilicities of different atoms. This atomic specific hydrophilicity scale measured by surface-bound water molecules was validated by

Received January 31, 1995; revision accepted June 2, 1995.
Address reprint requests to Leslie A. Kuhn, Department of Biochemistry, Michigan State University, East Lansing, MI 48824-1319 or to Elizabeth D. Getzoff, Department of Molecular Biology, The Scripps Research Institute, Department of Molecular Biology, 10666 North Torrey Pines Road, La Jolla, CA 92037.
Present address of Craig A. Swanson: Department of Chemistry, Box 39, Stanford University, Stanford, CA 94305.

the close correspondence of synthesized and observed specific hydrophilicity values for the 19 different amino acid side chains. Synthesized side-chain values were calculated by summing atomic hydrophilicity values for the atoms in each side chain. Thus, this new structure-based, experimental atomic scale is applicable for calculating hydrophilicity for functional groups similar to those in proteins, including discontinuous epitopes and drug molecules. Here, we use it to analyze complementarity in protein:protein and protein:drug interfaces.

METHODS

Developing a Crystallographic Hydration Database for Proteins

Specific hydrophilicity scales were determined from a database of 56 high-resolution crystallographic protein structures with reliable bound water positions by evaluating for each type of atom or residue the average number of closely bound water molecules with crystallographically defined positions. Only protein structures with resolution $\leq$2.1 Å, R-factor $\leq$23%, no interpenetrations of protein molecules in the crystal lattice, a range of temperature factors or occupancies indicating rigorous refinement, and >20 bound waters were included. Protein Data Bank codes for these proteins are 451c, 2act, 2alp, 4ape, 2apr, 2aza, 1bp2, 3c2c, 1ccr, 2ccy, 2cdv, 2cga, 5cha, 5cpa, 2cpp, 1ctf, 2cts, 2cyp, 4dfr, 3ebx, 1ecd, 3est, 1fb4, 1gcr, 1gd1, 1gp1, 3grs, 4hhb, 1ins, 2lhb, 3lym, 2lzm, 1mbd, 2mhr, 2ovo, 9pap, 1pcy, 2pka, 2prk, 5pti, 2rhe, 3rp2, 5rsa, 5rxn, 1sgc, 1sn3, 2sod, 2tgp, 2tgt, 3tln, 1ton, 1tpa, 1tpp, 1ubq, 3wga, and 2wrp.

Each water molecule within 4.0 Å (1 Å = 0.1 nm) of a surface-exposed protein atom was assigned to the single closest such atom, thus requiring that each hydrated protein atom be the primary contact for its bound water molecules. This assignment still allowed multiple water molecules to hydrate one protein atom. Only surface-exposed atoms and residues were considered, so that all were compared in the same solvent-exposed environment. For comparison, statistics were also gathered for assigning each water molecule to *all* surface-exposed protein atoms within 4 Å. In both cases, only nonhydrogen atoms were considered, because crystallographic data usually do not provide hydrogen positions. To avoid potential artifacts, water molecules within protein–protein interfaces (in oligomeric interfaces or crystal contacts) were not included, since their binding determinants are potentially different.

A 4 Å cut-off distance was used to include all water molecules in the first hydration shell, defined by the observed radial distribution of water molecules (see Fig. 1). The 4 Å distance also allowed for inclusion of all van der Waals contacts of water molecules with protein, some of which would have been excluded automatically by the 3.5 Å criterion used in other studies. More restrictive hydrogen-bond criteria were intentionally avoided, because this would introduce the assumption that hydrogen bonding is necessary for water binding (a bias already present in the assignment of many crystallographic water molecules). Protein surface shape is also a major determinant of water binding.[30] Geometric criteria for hydrogen bonds between water and protein led to the exclusion of ~40% of bound water molecules in another study.[10] While this could be advantageous for the purpose of predicting water molecule positions, our goal was to study all water molecules in the first hydration layer.

Measuring Specific Hydrophilicity by Water Binding

The interactions between 10,837 surface-bound water molecules and the atoms in 14,001 surface-exposed residues were compiled, and hydration values were calculated by dividing the total number of water molecules closely bound to side-chain and main-chain atoms in solvent-exposed residues by the total number of solvent-exposed residues. Residues with any exposed surface area[31] were considered solvent-exposed. Similarly, atomic hydrophilicity was determined from the number of water molecules closely bound to each type of surface-exposed atom (e.g., Ala carbonyl oxygen, Arg carbonyl oxygen, . . .) divided by the total number of surface-exposed atoms of that type. These specific hydrophilicity measures are equivalent to the average number of closely-bound water molecules per surface-exposed side chain, residue, or atom. Normalization by number of surface-exposed atoms rather than per unit surface area was chosen for two reasons. First, for many applications of hydrophobicity–hydrophilicity scales, the protein structure is not known, and surface area values are unavailable. Second, we investigated summing atomic hydration values normalized by surface area to synthesize side chains and found that these values do not approximate the observed side-chain hydration/area values well. This is likely because an extra assumption is being made: that each atom has average surface exposure. Hydration normalized by number of atom occurrences does approximate functional group values well, as shown in the Results.

Thanki et al.[8] have studied crystallographic hydration, with a focus on the spatial arrangement of bound water molecules around side chains. Their study includes both buried and exposed residues and all water molecules within a specified distance of the protein (3.5 Å for polar atoms and side chains, 5.0 Å for nonpolar) in 16 high-resolution protein structures, and uses the percentage of side and main chains with at least one bound water molecule as a measure of hydrophilicity. In their study as in ours, hydrogen-bond criteria are not imposed. In order to compare water binding of different residues within

the same environment, we considered only surface-exposed residues in contact with bulk solvent. For each protein atom or residue, we used the number of closely bound water molecules, not the proximity of at least one water molecule, as a hydrophilicity measure. We also compared the results of assigning each water molecule to all residues within the threshold distance, versus assigning each to the closest protein atom.

Mapping Atomic Specific Hydrophilicity Values Onto Molecular Interfaces

To analyze atomic hydrophilic complementarity at interfaces between protein surfaces, buried and solvent-accessible molecular surfaces[31] were calculated with a 1.6 Å radius probe and the following van der Waals radii including implicit hydrogens: O, 1.40 Å; OH, 1.60 Å; N, 1.54 Å; NH, 1.70 Å; NH_2, 1.80 Å; NH_3, 2.00 Å; CH, CH_2, CH_3, 2.00 Å; C, 1.74 Å; CH(sp^2), 1.86 Å; S, 1.80 Å; and SH, 1.85 Å. Surfaces were either calculated as dot surfaces with 20 dots per $Å^2$ by using *MS*[31] and visualized using Biosym's *InsightII* and the authors' *srftobiosym,* or calculated as triangulated surfaces by using *MSP*[32] and visualized using *AVS*[33] and *AVS* modules developed by Michael Pique and colleagues at The Scripps Research Institute.

Two methods were developed for evaluating atomic hydrophilicity at protein surfaces. The first method maps atomic hydrophilicity values directly onto the surface of each atom. The specific hydrophilicity value for one of five atomic groups (see Table III) is assigned to each surface-exposed atom, and each molecular surface dot (or triangle vertex) is color-coded according to the assigned value of the nearest atom. The second method evaluates the average hydrophilicity of an atom's neighborhood, and was motivated by the observation that any protein atom contacting the inhibitor is likely to contact more than one inhibitor atom (or vice versa). The atomic neighborhood hydrophilicity of a surface point is calculated by averaging the atomic specific hydrophilicity values (see Table III) for all surface-exposed atoms from the same molecule that are within 4 Å of that surface point. These atomic neighborhood hydrophilicity values are color-mapped to the surface points (vertices of triangles), and the colors are smoothed across each triangular surface face by *AVS*.

RESULTS AND DISCUSSION

Side-Chain Specific Hydrophilicity Correlates With Hydrogen-Bonding Capacity

The specific hydrophilicity scale (Table I) measured by side-chain hydration discriminated between hydrophilic and hydrophobic residues and linked hydration to hydrogen-bonding capacity. The observed radial distribution function for water molecules within our database (Fig. 1) was used to define the water molecules within the first hydration shell to be included in determining the hydrophilicity scales. The number of bound water molecules as a function of the distance from the nearest protein atom falls off beyond 4 Å (Fig. 1). This matches water radial distribution functions published for insulin[34] and liquid water,[35] which show the second hydration shell to begin at 4 Å.

Specific hydrophilicity (black bars, Fig. 2) varied 10-fold among the different side chains, from an average of ~1.2 hydrations per occurrence for Arg down to ~0.12 for Val. The hydrophilic side chains Arg, Glu, Asp, Gln, Asn, Lys, His, Tyr, Thr, and Ser each bound more than 0.6 water molecules on average, while the more hydrophobic side chains Trp, Pro, Met, Ala, Phe, Leu, Ile, Cys, and Val each bound fewer than 0.4. This signifies that bound water molecules can have a nonpolar side chain as their primary (closest) contact. Water binding data also provided independent hydrophilicity values for the different residues' main chains (Table I). Most main chains bind an average of ~0.4 water molecules; however, the main chains of the smallest residues were significantly more hydrated on average (Ala, 0.58 hydrations, and Gly, 0.68 hydrations), likely due to their structural context and greater accessibility of the main chain.

Specific hydrophilicity for side chains correlated highly (linear correlation coefficient, $r = 0.97$) with the number of potential unbifurcated hydrogen bonds (according to Stickle et al.[36]) the side chains could form: 5 for Arg, 4 for Glu, Asp, Gln, Asn; 3 for Lys, Thr, Ser; 2 for His, Tyr; 1 for Trp; and 0 for the other side chains. Other studies cite hydrogen bonding as a major determinant for water affinity[1,11,12] and show aromaticity to be favorable for hydrogen bonding.[13,37] This agrees with our side-chain hydration data ranking His and Tyr side chains as slightly more hydrophilic than Thr and Ser, despite having more carbon atoms and less hydrogen-bonding capacity. Complementarily, Fourier transform infrared spectroscopy establishes that hydrogen bonding *between water molecules* within the hydration shells of free amino acids increases with increasing side-chain hydrophobicity.[38] In folded proteins, hydration through protein–water hydrogen bonds competes not only with this interwater hydrogen bonding, but also with intraprotein hydrogen bonding, as discussed later in the context of atomic hydrophilicity.

Comparison of Side-Chain Specific Hydrophilicity With Other Scales

When compared with other hydrophilicity scales, the specific hydrophilicity scales (Table I) correlated best with experimental, rather than than derived, measures of partitioning between aqueous and organic phases. The linear correlation coefficient with side-chain free energy of transfer between cyclohexane and water[39,40] was $r = -0.95$ (Fig. 3) and with

TABLE I. Specific Hydrophilicity Scales for Main Chains, Side Chains, and Residues in Folded Proteins*

	Surface-exposed occurrences	Hydrations per occurrence		
		Main chain	Side chain	Residue
ALA	947	0.58	0.20	0.78
ARG	456	0.42	1.16	1.58
ASN	660	0.37	0.83	1.20
ASP	660	0.38	0.97	1.35
CYS	295	0.42	0.12	0.55
GLN	452	0.35	0.84	1.19
GLU	643	0.45	1.00	1.45
GLY	1040	0.68	—	0.68
HIS	299	0.28	0.71	0.99
ILE	431	0.33	0.14	0.47
LEU	692	0.39	0.16	0.56
LYS	837	0.35	0.75	1.10
MET	151	0.39	0.27	0.66
PHE	357	0.30	0.17	0.47
PRO	580	0.35	0.34	0.69
SER	963	0.39	0.61	1.00
THR	775	0.42	0.64	1.05
TRP	186	0.34	0.37	0.70
TYR	455	0.34	0.66	1.00
VAL	687	0.38	0.12	0.51

*The average number of hydrations (bound water molecules) per occurrence measures specific hydrophilicity and was calculated by dividing the total number of water molecules bound to all surface-exposed main chains, side chains, or residues by the total number of surface-exposed occurrences of the residue. Of all residues in the 56-protein database, 82.6% had some surface exposure, as gauged by using the *MS* solvent-accessible molecular surface program[31] with a probe radius of 1.6 Å, density of 4 dots per Å^2, and van der Waals radii given in the text.

residue retention on reversed-phase HPLC, using a pH 7.5 average scale,[28] was $r = -0.93$. (Comparisons to published side-chain hydrophilicity scales were made with our side-chain scale; likewise, residue scales were compared to our residue scale.) Correlation was also high ($r = 0.95$) between side-chain specific hydrophilicity and the percentage, in 16 crystallographic structures, of exposed and buried side chains having a water molecule[8] within 3.5 Å (see Table 4 in ref. 8). Slightly lower correlations were found with consensus residue scales determined by averaging several different experimental measures (correlation with consensus hydrophobicity,[41] $r = -0.91$; with hydropathy,[42] $r = -0.89$); with an octanol–water partition scale,[29] r −0.86; and with an area-based scale,[22] $r = -0.80$. No consistent outliers were found when the side-chain or residue-specific hydrophilicity scales were compared to a number of hydrophilicity scales.

The scales described above differ in how residues are represented (residue within a folded protein versus amino acid analog or peptide) and in the property being measured (tendency of water to bind one side chain over another versus tendency of side chain to prefer water over another solvent), as well as reflect intrinsic variability and errors in the measurements. The scales presented here measure specific hydrophilicity as the propensity of surface-exposed atoms and residues in folded proteins to bind water molecules, and thus may contribute to a more complete description of protein solvation that includes terms for both explicit binding to solvent and extent of exposure to bulk solvent. Note that even though crystallographic hydration represents only the well-ordered water molecules in contact with protein, this scale correlates highly with free energy hydrophilicity measures. Thus, our results suggest that interactions between protein and bound water are fundamentally similar to the interactions between protein and bulk (fluid) water, which is measured in the partition experiments.

Atomic Hydrophilicity Values Cluster

Since several earlier studies assumed that certain atom types had equivalent hydrophilicity and then derived group hydrophilicity values, a major goal of this work was to measure the hydrophilicity of each atom type (presented in Table II), then deduce which atoms were equivalently hydrophilic. When the average number of hydrations per occurrence was graphed for each atom type (Fig. 4), specific hydrophilicity values for atom types with similar chemistry were found to cluster. Neutral and negatively charged oxygen atoms were the most hydrophilic, followed by positively charged nitrogen atoms, followed by neutral nitrogens, followed by carbons and sulfurs ($O \sim O^- > N^+ > N \gg C \sim S$, where "$\sim$" indicates similar hydrophilicity values, ">" means "more hydrophilic than," and $\gg$ means "much more hydrophilic than"). Specific hydrophilicity values for charged and neutral nitrogen and oxygen atoms were sufficiently different to keep them as separate groups, and carbon and sulfur atoms were sufficiently similar to be grouped together (Table III).

Five atomic groups were derived: (1) neutral oxygen group: all carbonyl oxygen atoms plus Gln, Asn, Ser, Thr, and Tyr side-chain oxygens; (2) negative oxygen group: Asp and Glu side-chain oxygens; (3) positive nitrogen group: His, Lys, and Arg side chain nitrogens excepting Arg N_ϵ, which is fairly neutral (according to partial charge values tabulated in Biosym Technologies' Discover consistent valence forcefield[43]); (4) neutral nitrogen group: all main-chain nitrogens, Trp, Asn, and Gln side-chain nitrogens, and Arg N_ϵ; and (5) carbon/sulfur group: all carbon and sulfur atoms. Changing the criteria for membership in the positive and neutral nitrogen groups (from N^+ including all Arg, His, and Lys side-chain nitrogen atoms to N^+ including only Arg $N_{\eta 1}$, Arg $N_{\eta 2}$, and Lys N_ζ) had a negligible effect on the value for neutral nitrogen hydration (excluding the Pro main-chain nitrogen, values for neutral nitrogen atom types ranged from 0.353–0.358 hydrations/oc-

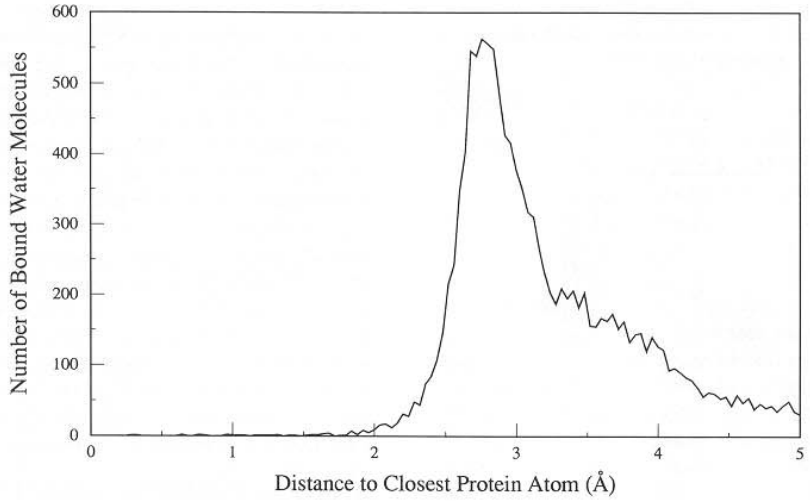

Fig. 1. Radial distribution function for all water molecules in the database's 56 protein structures. The number of water molecules is graphed as a function of distance between the oxygen atom of each water molecule and the nearest protein atom. The peak at ~2.8 Å encompasses water molecules that are hydrogen-bonded to protein. Water molecules up to 4 Å from protein surface atoms, comprising the first hydration shell of proteins, were included in the atomic and residue hydrophilicity study. These water molecules interact with protein via hydrogen bonds, shape-dependent constraints on water mobility, packing, or electrostatic interactions.

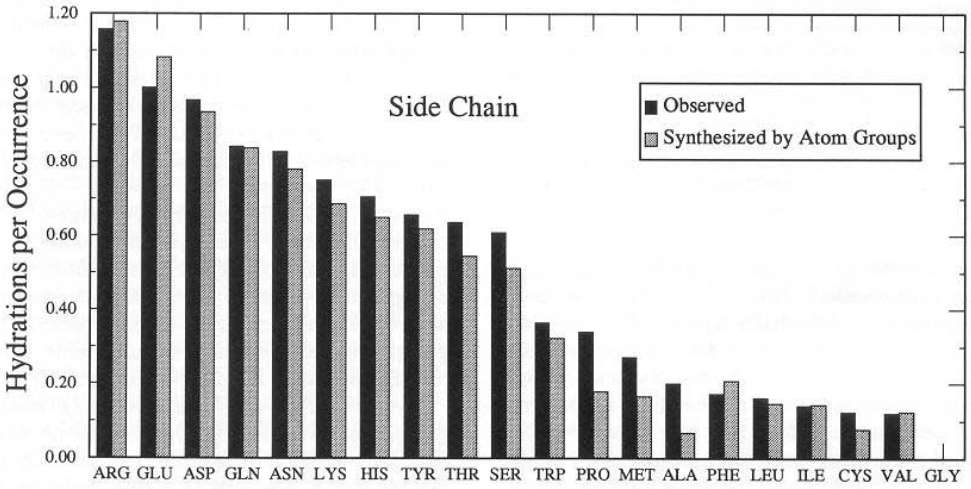

Fig. 2. Side-chain specific hydrophilicity at protein surfaces. Specific hydrophilicity was measured by the average number of bound water molecules (black bars) and also synthesized (gray bars) from atomic values as described in the text. The most hydrophilic side chains (Arg through Trp) all contained oxygen and nitrogen atoms with hydrogen-bonding capability, whereas side chains containing only carbon, sulfur, and hydrogen (Pro through Val) bound significantly less water.

currence) and limited effect on positive nitrogen hydration values (0.420–0.460 hydrations/occurrence). In each of the two negatively charged side chains, the two oxygen atoms had similar average hydration values (Asp OD1 and OD2, 0.546 hydrations/occurrence each; Glu OE1, 0.481, and OE2, 0.469). If we reassigned the Asp atom names OD1 and OD2 with the new OD1 consistently being the side-chain oxygen atom with more surface exposure (and similarly reassigned Glu OE1 and OE2), we found that new OD1 was more exposed and more hydrated than new OD2, while new OE1 was more exposed but somewhat less hydrated than new OE2. Thus, there was no obvious trend in increased hydration with

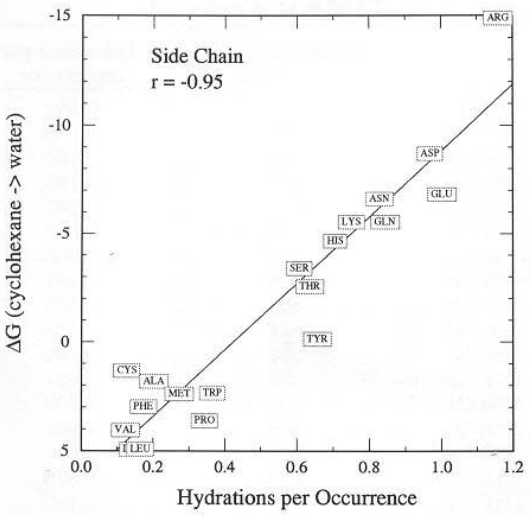

Fig. 3. Comparison between two experimentally determined side-chain hydrophilicity measures: free energy of transfer between a nonpolar phase (cyclohexane) and water, and specific hydrophilicity measured by water binding. Free energy of transfer[39,40] and average number of hydrations were linearly related, with a correlation coefficient of $r = -0.95$. (Hydrophobicity scales are inversely correlated to hydrophilicity scales, and therefore the comparison yields a negative correlation coefficient.) Whereas Arg appears to be less hydrated than expected by comparison to the cyclohexane→water partition scale, Arg is more hydrated than expected when compared to an HPLC-based scale.[28] The relative hydration of residues can be understood by examining the hydrogen-bonding atoms' environments, as discussed in the atomic hydrophilicity section.

increased surface exposure. Furthermore, all negatively charged oxygen atoms except OD1 remained significantly less hydrated than neutral oxygen atoms.

Assigning each bound water molecule to the closest atom (as presented in the figures and tables), rather than to all atoms within the 4.0 Å cut-off distance, improved discrimination between atom types. When we compared hydration values for each chemically equivalent atom type (e.g., all methyl carbons, or all carbonyl carbons) derived from assigning each bound water molecule to the closest protein atom, versus assigning each water molecule to all protein atoms within 4.0 Å, we found that only in the case of closest assignment were chemically equivalent atoms hydrated to a similar extent. For instance, when total hydration per occurrence was measured, carbon atoms had similar values to oxygen and nitrogen atoms. When closest hydration was measured, carbon atoms were only ~1/10 as hydrated as polar atoms, as expected. Furthermore, changing from measuring closest hydration to total hydration resulted in roughly the same proportionality constant for most polar atoms, while there was no such linear relationship for carbon atoms. Both of these observations pointed to total hydration as being too inclusive and nonspecific, possibly because

TABLE II. Specific Hydrophilicity Values for All Atom Types in Proteins*

	Surface-exposed occurrences	Hydrations per occurrence
ALA N	375	0.355
ALA CA	543	0.072
ALA C	305	0.007
ALA O	635	0.584
ALA CB	819	0.232
ARG N	115	0.470
ARG CA	179	0.039
ARG C	86	0.012
ARG O	216	0.606
ARG CB	275	0.033
ARG CG	283	0.085
ARG CD	337	0.172
ARG NE	299	0.237
ARG CZ	230	0.026
ARG NH1	384	0.440
ARG NH2	398	0.445
ASN N	177	0.367
ASN CA	329	0.021
ASN C	181	0.017
ASN O	362	0.448
ASN CB	500	0.088
ASN CG	327	0.040
ASN OD1	484	0.459
ASN ND2	560	0.464
ASP N	229	0.301
ASP CA	370	0.011
ASP C	157	0.006
ASP O	352	0.503
ASP CB	522	0.082
ASP CG	463	0.037
ASP OD1	522	0.546
ASP OD2	538	0.546
CYS N	75	0.400
CYS CA	110	0.036
CYS C	69	0.014
CYS O	151	0.583
CYS CB	160	0.100
CYS SG	151	0.139
GLN N	82	0.415
GLN CA	200	0.040
GLN C	114	0.009
GLN O	227	0.502
GLN CB	296	0.064
GLN CG	310	0.116
GLN CD	255	0.020
GLN OE1	351	0.453
GLN NE2	365	0.416
GLU N	213	0.343
GLU CA	348	0.040
GLU C	163	0.006
GLU O	392	0.513
GLU CB	497	0.070
GLU CG	532	0.109
GLU CD	510	0.014
GLU OE1	574	0.481
GLU OE2	552	0.469
GLY N	555	0.312
GLY CA	954	0.139

(continued)

TABLE II. (Continued)

	Surface-exposed occurrences	Hydrations per occurrence
GLY C	542	0.028
GLY O	725	0.521
HIS N	64	0.359
HIS CA	129	0.023
HIS C	45	0.022
HIS O	129	0.442
HIS CB	162	0.099
HIS CG	67	0.060
HIS CD2	171	0.140
HIS ND1	145	0.441
HIS CE1	224	0.165
HIS NE2	187	0.353
ILE N	79	0.380
ILE CA	124	0.024
ILE C	43	0.000
ILE O	202	0.545
ILE CB	112	0.018
ILE CG1	194	0.031
ILE CG2	241	0.124
ILE CD1	264	0.087
LEU N	159	0.384
LEU CA	193	0.010
LEU C	130	0.000
LEU O	367	0.548
LEU CB	315	0.035
LEU CG	198	0.020
LEU CD1	383	0.125
LEU CD2	426	0.117
LYS N	207	0.271
LYS CA	425	0.012
LYS C	213	0.009
LYS O	468	0.474
LYS CB	616	0.060
LYS CG	655	0.076
LYS CD	703	0.081
LYS CE	766	0.129
LYS NZ	811	0.476
MET N	35	0.486
MET CA	66	0.030
MET C	33	0.030
MET O	66	0.591
MET CB	76	0.132
MET CG	95	0.095
MET SD	65	0.015
MET CE	101	0.208
PHE N	69	0.406
PHE CA	142	0.014
PHE C	54	0.000
PHE O	153	0.503
PHE CB	172	0.029
PHE CG	26	0.000
PHE CD1	120	0.067
PHE CD2	157	0.076
PHE CE1	142	0.099
PHE CE2	187	0.048
PHE CZ	173	0.081
PRO N	63	0.016
PRO CA	285	0.049
PRO C	172	0.035
PRO O	369	0.488
PRO CB	455	0.132
PRO CG	452	0.164
PRO CD	437	0.146
SER N	372	0.245
SER CA	581	0.022
SER C	276	0.000
SER O	558	0.491
SER CB	828	0.150
SER OG	814	0.568
THR N	212	0.292
THR CA	386	0.021
THR C	154	0.000
THR O	419	0.601
THR CB	501	0.054
THR OG1	629	0.536
THR CG2	669	0.193
TRP N	42	0.429
TRP CA	65	0.046
TRP C	32	0.031
TRP O	72	0.569
TRP CB	81	0.049
TRP CG	5	0.000
TRP CD1	87	0.149
TRP CD2	15	0.067
TRP NE1	86	0.349
TRP CE2	20	0.100
TRP CE3	37	0.027
TRP CZ2	83	0.072
TRP CZ3	56	0.071
TRP CH2	61	0.115
TYR N	99	0.374
TYR CA	155	0.006
TYR C	68	0.074
TYR O	198	0.551
TYR CB	216	0.069
TYR CG	46	0.022
TYR CD1	168	0.083
TYR CD2	178	0.051
TYR CE1	250	0.068
TYR CE2	262	0.050
TYR CZ	111	0.036
TYR OH	356	0.635
VAL N	166	0.482
VAL CA	226	0.013
VAL C	84	0.000
VAL O	339	0.534
VAL CB	232	0.000
VAL CG1	459	0.085
VAL CG2	416	0.106

*The average number of hydrations (bound water molecules) per occurrence measures specific hydrophilicity and was calculated by dividing the total number of water molecules bound closely to all surface-exposed atoms by the total number of surface-exposed occurrences of the atom.

(*continued*)

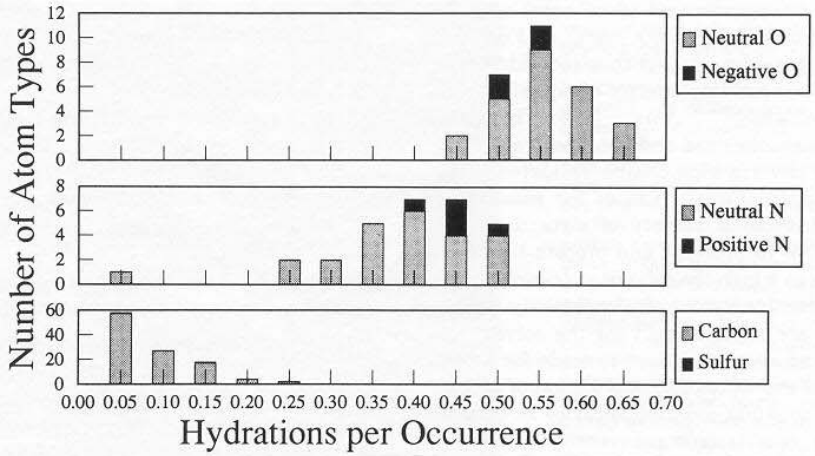

Fig. 4. Distribution of atomic hydration values. To determine which atoms are similar or distinct with respect to water binding, we graphed the number of atom types (e.g, Ala amide nitrogen, Ala C_{α}, . . .) at each hydration per occurrence value. Each atom type contributed one vertical unit to the graph, and atoms were grouped as described in the text. Oxygen atoms were the most hydrated (top graph), with negatively charged oxygen (black bars) slightly less hydrated on average than neutral oxygen (gray bars). Nitrogens (center graph) were the next most hydrated, overlapping the oxygen distribution, and positively charged nitrogens (black bars) were somewhat more hydrated than neutral nitrogens (gray bars). Proline's amide nitrogen, with no hydrogen-bonding capacity, had the lowest nitrogen hydration value (leftmost bar). Carbon and sulfur atoms (bottom graph; note change of *y*-axis scale) were the least hydrated, with sulfur values at 0.05 and 0.15 hydrations per occurrence.

TABLE III. Specific Hydrophilicity Values for Protein Atoms

	Hydrations per occurrence*
Neutral oxygen	0.53
Negative oxygen	0.51
Positive nitrogen	0.44
Neutral nitrogen	0.35
Carbon, sulfur	0.08

*The average number of hydrations per occurrence was calculated over all atoms within each group (see text).

water molecules were also being assigned to carbon atoms proximal to hydrophilic atoms.

Synthesizing Specific Hydrophilicity Values for Functional Groups

To evaluate whether these five specific hydrophilicity values for atoms (Table III), which are averages over many atoms in many residues, can be used to represent hydrophilicity in different contexts, atomic hydrophilicity values were summed for the atoms in each of the 19 side-chain types to synthesize side-chain values. Synthesizing atomic hydrophilicity values to calculate the hydrophilicity of functional groups requires knowledge, or an estimate, of the surface exposure of each atom. This requirement is based both upon the assumption that interaction with water requires surface exposure, and upon our tests showing that when surface exposure is ignored, the synthesized hydrophilicity values are consistently overestimated. However, a binary measure of surface exposure—either an atom is exposed and thus can bind water, or it is not exposed and cannot bind water—was sufficient for very good synthesized estimates of side-chain hydrophilicity (see Fig. 2). We were unable to deduce any simple set of atomic groups (e.g., neutral O, negatively charged O, or other groupings based upon covalent bond partners and branching) that would give good approximations of both the likelihood of being surface exposed, and the average number of bound water molecules; atom types with similar likelihood of being surface-exposed do not group the same way as atom types with similar hydration. However, for an atom of unknown surface exposure, the likelihood of surface exposure can be calculated as follows. The probability that an Asp $O_{\delta 1}$ atom is surface-exposed is simply (number of surface-exposed Asp $O_{\delta 1}$ atoms in the database)/(number of Asp residues in the database), or 522/660 (values from Tables I and II). An estimate of the hydration of an Asp $O_{\delta 1}$ atom of unknown surface exposure equals its likelihood of being surface-exposed multiplied by the average number of waters bound to a surface-exposed negatively charged oxygen atom (found in Table III), or (522/660) $\times$ 0.51 hydrations = 0.40 hydrations on average.

Good estimates of the observed side-chain hydrophilicity values were obtained using this synthesis procedure, as shown in Figure 2. Group atomic hydrophilicity values in Table III were multiplied by the estimates of the atoms' surface exposures from the values in Tables I and II. The correlation coeffi-

cient between the synthesized (grey bars) and observed side-chain values (black bars) in Figure 2 was 0.99, and between the synthesized side-chain values and side-chain free energy of transfer from cyclohexane to water[39,40] was −0.94. The close match between synthesized and observed values for the chemically diverse side chains validates using these atomic hydrophilicity values for measuring the hydrophilicity of a variety of structural and functional groups in proteins and protein-like molecules. In cases in which the solvent exposures of the molecule's atoms are known, hydrophilicity values from Table III are summed only for the solvent-exposed atoms, and no contribution is made for atoms without solvent exposure. The application of atomic hydrophilicity to study complementarity in protein interfaces (e.g., the trypsin:pancreatic trypsin inhibitor and HIV protease:U-75875 interfaces discussed later in this paper) involves surface atoms only; thus, the values in Table III can be used directly, without requiring an estimate of surface exposure.

Comparing Atomic Specific Hydrophilicity With Solvation Models

Rather than measure hydrophilicity from water binding, other researchers have developed atomic hydrophilicity scales by deriving best fits to transfer free energy or solvent-exposed surface area measurements for residues or their fragments. Holm and Sander[26] model the solvation preference of atoms in proteins by evaluating the percent of each atom's neighbors that are protein, modeling the rest of space as solvent, then calculating the likelihood that each atom type will have a given percentage of solvent neighbors. (The Holm and Sander solvation preference value for an atom depends on the atom's degree of accessibility; therefore, there is no direct way to compare solvation preference to specific hydrophilicity.) Eisenberg and colleagues[23,24] model an atom's free energy of transfer from the protein interior into aqueous solvent by the product of the atom's accessible surface area and a solvation parameter specific to that atom group, which is derived by fitting the sum of atomic values for each residue to experimentally measured transfer free energy values for residues[29]. Their order in solvation parameters is $N^+ \sim O^- > O/N > S > C$, where O/N indicates that they assumed O and N to be equivalent and grouped them before calculating the solvation parameter.

The scale of Lesser and Rose[44] models hydrophilicity by mean atomic surface area buried upon folding, yielding the hydrophilicity order $O^-/N^+ > O/N > C/S$. Abraham and Leo[25] use partition coefficients measured for many organic compounds to derive free energy of transfer values for amino acid fragments, resulting in a ranking of $COO^- > NH_3^+ > OH \sim NH_2 > S > C$ (see their Table 1). Thanki et al.[8] tabulate the total number of water molecules within 3.5 Å of polar functional groups in their study of the stereochemistry of protein–water interactions. When we divided their total number of water molecules per polar group by the total number of residues containing that group (their Table 3), the derived hydrophilicity order was $O^- > N^+ > O > N$.

Probing the Influence of Structural Environment on Atomic Hydrophilicity

Our experimentally defined specific hydrophilicity scale agrees with the consensus of the above scales, $O^- \geq N^+ > S \geq C$. To understand why we find neutral oxygen to be hydrated similarly to negatively charged oxygen (Table III), the neighborhood of each oxygen or nitrogen atom was analyzed with regard to how many hydrogen bonds the atom could potentially form, how this number related to the number of water molecules bound, and what balance was attained in the competition between hydrogen bonds to water molecules and to protein atoms. For surface-exposed oxygen and nitrogen atoms, the percentage of potential hydrogen bonds to water molecules (out of all possible hydrogen bonds) was estimated by dividing the number of neighboring water molecules (within 4 Å) by the number of neighboring water molecules and protein atoms with appropriate hydrogen-bond (or salt-bridge) character. By multiplying this expected percentage of hydrogen bonds to water by the maximum number of unshared hydrogen bonds each atom can make (e.g., a carbonyl oxygen can make 2 hydrogen bonds; an amide nitrogen, one), the following numbers of hydrogen bonds to water were obtained: 1.53 per neutral oxygen atom, 1.40 per positive nitrogen atom, 1.38 per negative oxygen atom, and 0.79 per neutral nitrogen atom. This ranking in number of hydrogen bonds to water ($O > N^+ \sim O^- > N$) explains why the atomic specific hydrophilicity scale ranks $O \geq O^-$ and $N^+ \geq N$. The greater hydration of O^- relative to N^+ in the atomic hydrophilicity scale is consistent with the stricter geometrical constraints on hydrogen-bond angles for nitrogen atoms than for oxygen atoms.[11] The similarity in hydrogen-bond and specific hydrophilicity ranking for atoms supports hydrogen bonding (and competition for hydrogen bonds by nearby protein atoms) as a major determinant in water binding by protein atoms.

Atomic Electrostatic Forces Influence Specific Hydrophilicity

Both electronegativity and charge delocalization, in addition to hydrogen-bonding geometry, also tend to make nitrogen atoms less hydrated than oxygen atoms, consistent with specific hydrophilicity data. Electronegativity makes O–H bonds more polar

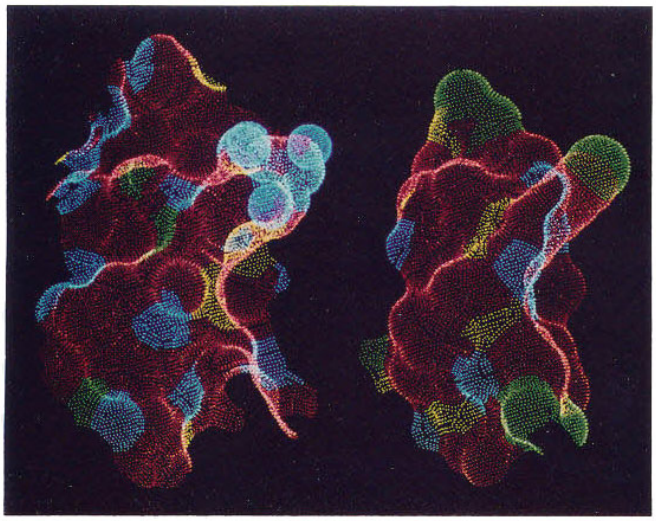

Fig. 5. Evaluating molecular recognition using atomic hydrophilicity. The molecular surfaces of bovine trypsin (left) and BPTI (right) buried in their interface have been shifted apart sideways to allow comparison of the surfaces' specific hydrophilicities. The inhibitory lysine of BPTI (right) protrudes from the center of the surface and is encompassed by the active site pocket of trypsin (left). Solvent-accessible molecular surface points were color-coded by the specific hydrophilicity values of atoms contributing to the surface, with a gradient from most hydrophilic (blue) to least (red): neutral oxygen, blue; negative oxygen, blue-green; positive nitrogen, green; neutral nitrogen, yellow; carbon and sulfur, red. Note patches of similar surface coloration between the trypsin and inhibitor surfaces, indicating hydrophilic complementarity at the atomic scale. The hydrophobic patch in the central pocket of trypsin's interface surface is offset by three buried water molecules (blue spheres in left figure), one making a critical contact with the lysine side chain of the inhibitor. These water molecules also contribute to favorable interactions with the lysine by providing hydrophilic contacts and filling small cavities in the interface. See ref. 49 for a detailed analysis of water binding to trypsin and its active site. Trypsin and BPTI structural coordinates are from Protein Data Bank entry 1TPA,[45–47] and buried surfaces were calculated using *MS*[31] as described in Methods.

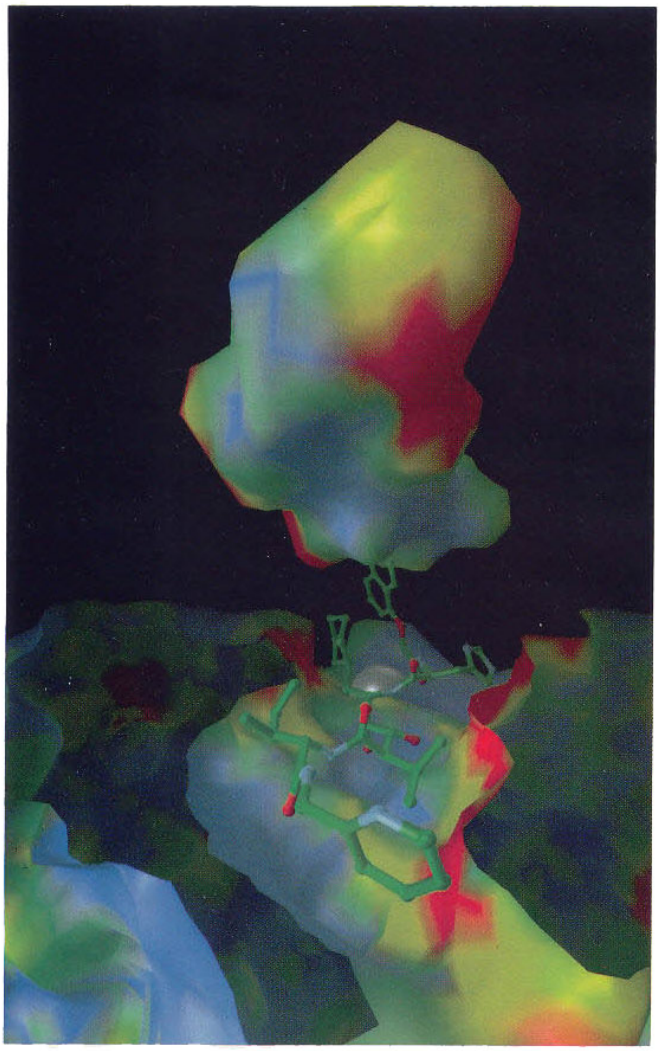

Fig. 6. Complementarity of atomic neighborhood hydrophilicity between the HIV-1 protease and a high-affinity inhibitor. The solvent-accessible molecular surface[31] of HIV-1 protease (Protein Data Bank entry 1HIV) appears at bottom, sliced along the long axis (from front of view to back) of its cylindrical active site, with the elongated surface of its semi-peptidyl inhibitor, U-75875 or Noa-His-HchΨ [CH(OH)CH(OH)] Vam-Ile-Amp,[48] shifted above the ball-and-stick representation of its atoms (carbon, green; nitrogen, blue; oxygen, red). The white sphere in the background is Water 303, which binds to the main chain of the naphthoxyacetyl (Noa) residue in the inhibitor. Atomic neighborhood hydrophilicity values, reflecting the hydrophilic neighborhood that an atom on the adjacent molecule senses upon binding, were calculated as described in Methods. Values for the protease and inhibitor were color-mapped to distinguish peaks in the observed distribution of atomic neighborhood hydrophilicity values (0–0.5 hydrations, red; 0.5–0.9 hydrations, yellow; 0.9–1.4 hydrations, green; >1.4 hydrations, blue). Hydrophilic complementarity is apparent between the backbone of the U-75875 inhibitor (blue regions along long axis of the inhibitor surface at top) and the active-site groove of the HIV-1 protease (blue patches on the protease surface beneath the inhibitor ball-and-stick model) and bound Water 303. Complementarity of hydrophobic regions (yellow and red) is seen between the inhibitor side-chains and the active-site wall, at right.

than N–H bonds. Therefore, oxygen atoms tend to form somewhat stronger hydrogen bonds than nitrogen atoms, and thus be more hydrated than nitrogen atoms. Charge delocalization from nitrogen, but not oxygen, atoms onto adjacent carbon atoms[25] also influences the atomic hydration values, tending to lessen the hydrophilicity of nitrogen atoms, but not oxygen atoms. With the exception of branching carbon atoms (which were consistently observed to have reduced hydration regardless of covalent neighbors), carbon atoms bonded to positively charged nitrogen atoms were ~2 times as hydrated as expected from their group values; enhanced hydration of carbons atoms bonded to negatively charged oxygen was neither observed nor expected.

Analyzing Hydrophilic/Hydrophobic Complementarity at Molecular Interfaces

Just as the protein structural context gives insights into the chemical and structural factors affecting hydration, hydrophilicity values based on hydration can provide insights into molecular interactions at the atomic level. In protein–protein recognition, both shape and chemistry establish complementarity. Coloring the interacting surfaces of bovine pancreatic trypsin (EC 3.4.21.4) and bovine pancreatic trypsin inhibitor (BPTI), Protein Data Bank entry 1TPA,[45–47] by atomic specific hydrophilicity values (Table III) reveals hydrophilic and hydrophobic complementarity throughout the interface (Fig. 5). Three-fourths of all hydrophilic regions in each of the interfaces complement a hydrophilic region on the other molecule, and two-thirds of all hydrophilic regions show potential for complementary hydrogen bonds across the interface.

This method of analyzing hydrophilic/hydrophobic surface complementarity, which also has been used by others,[24] shows very local surface complementarity and is essentially equivalent to coloring surface patches by atom type. An alternative, and possibly more realistic, technique is to consider several neighboring atoms' contributions to the hydrophilic/hydrophobic contact; this mirrors the fact that any atom in the inhibitor is likely to contact several protein atoms simultaneously. We show this atomic neighborhood hydrophilicity in Figure 6 for the interface between HIV-1 protease and the semi-peptidyl U-75875 inhibitor (Protein Data Bank entry 1HIV)[48] by calculating, at each molecular surface point buried in their interface, the average atomic hydrophilicity within a 4 Å sphere. By applying the resultant atomic neighborhood hydrophilicity values as a color code on the molecular surfaces, the strong hydrophilic complementarity between the backbone of the inhibitor and the long groove axis of the protease active site, as well as the hydrophobic complementarity along the inhibitor's periphery, are evident (Fig. 6).

CONCLUSIONS

Specific hydrophilicity scales based on the binding of water molecules to surface-exposed protein atoms provide an experimental measure of water affinity for any atom, functional group, or residue in the context of the folded protein. The direct relationship between this atomic hydrophilicity scale and its structural database is advantageous because it allows the roles of specific chemical and structural factors in protein–water interactions, including competition for hydrogen bonds, to be analyzed. The appropriateness and applicability of the derived atomic groupings were validated by close similarity between the observed side-chain and residue specific hydrophilicity scales and those synthesized from atomic specific hydrophilicity values. For assessment of water–protein interactions and hydrophobic contacts relevant to a range of problems in drug design and molecular recognition, these structure-based, experimentally defined hydrophilicity scales may prove especially useful for comparison to hydrophilicity/hydrophobicity measured by bulk hydration.

ACKNOWLEDGMENTS

We thank Michael E. Siani and Gary H. Liao for software contributions and Lee Walters, Duncan McRee, Donald Bashford, and Brian Matthews for valuable discussions. This research was initially supported by American Cancer Society, California Division senior postdoctoral fellowship S65-92 to L.A.K., and by National Science Foundation grant DIR-8822385 to E.D.G., M.E.P., and J.A.T..

REFERENCES

1. Edsall, J. T., McKenzie, H. A. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. Adv. Biophys. 16:53–183, 1983.
2. Kuntz, I. D., Kauzmann, W. Hydration of proteins and polypeptides. Adv. Prot. Chem. 28:239–345, 1974.
3. Franks, F. "Water." London: The Royal Society of Chemistry. 1984.
4. Rupley, J. A., Careri, G. Protein hydration and function. Adv. Prot. Chem. 41:38–173, 1991.
5. Erickson, J., Neidhart, D. J., VanDrie, J., Kempf, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D. A., Knigge, M. Design, activity, and 2.8 Å crystal structure of a C_2 symmetric inhibitor complexed to HIV-1 protease. Science 249:527–533, 1990.
6. Schellman, J. A. A simple model for solvation in mixed solvents. Applications to the stabilization and destabilization of macromolecular structures. Biophys. Chem. 37: 121–140, 1990.
7. Robinson, C. R., Sligar, S. G. Molecular recognition mediated by bound water: A mechanism for star activity of the restriction endonuclease *EcoRI*. J. Mol. Biol. 234:302–306, 1993.
8. Thanki, N., Thornton, J. M., Goodfellow, J. M. Distributions of water around amino acid residues in proteins. J. Mol. Biol. 202:637–657, 1988.
9. Walshaw, J., Goodfellow, J. M. Distribution of solvent molecules around apolar side-chains in protein crystals. J. Mol. Biol. 231:392–414, 1993.
10. Roe, S. M., Teeter, M. M. Patterns for prediction of hydra-

tion around polar residues in proteins. J. Mol. Biol. 229: 419–427, 1993.
11. Baker, E. N., Hubbard, R. E. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44:97–179, 1984.
12. Jeffrey, G. A., Saenger, W. "Hydrogen Bonding in Biological Structures." New York: Springer-Verlag, 1991.
13. Levitt, M., Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. J. Mol. Biol. 201:751–754, 1988.
14. Otting, G., Liepinsh, E., Wüthrich, K. Protein hydration in aqueous solution. Science 254:974–980, 1991.
15. Clore, G. M., Bax, A., Wingfield, P. T., Gronenborn, A. M. Identification and localization of bound internal water in the solution structure of interleukin 1β by heteronuclear three-dimensional ^{1}H rotating-frame Overhauser ^{15}N-^{1}H multiple quantum coherence NMR spectroscopy. Biochemistry 29:5671–5676, 1990.
16. Levitt, M., Park, B. H. Water: now you see it, now you don't. Structure 1:223–226, 1993.
17. Matthews B. W. Solvent content of protein crystals. J. Mol. Biol. 33:491–497, 1968.
18. Carter, D. C., He, X.-M., Munson, S. H., Twigg, P. D., Gernert, K. M., Broom, M. B., Miller, T. Y. Three-dimensional structure of human serum albumin. Science 244: 1195–1198, 1989.
19. Zhang, X.-J., Matthews, B. W. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. Protein Sci. 3:1031–1039, 1994.
20. Chothia, C. Hydrophobic binding and accessible surface area in proteins. Nature (London) 248:338–339, 1974.
21. Richards, F. M. Areas, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. 6:151–176, 1977.
22. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. Science 229:834–838, 1985.
23. Eisenberg, D., McLachlan, A. D. Solvation energy in protein folding and binding. Nature (London) 319:199–203, 1986.
24. Eisenberg, D., Wesson, M., Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. Chem. Scripta 29A:217–221, 1989.
25. Abraham, D.J., Leo, A. J. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. Proteins 2:130–152, 1987.
26. Holm, L., Sander, C. Evaluation of protein models by atomic solvation preference. J. Mol. Biol. 225:93–105, 1992.
27. Nozaki, Y., Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. J. Biol. Chem. 246:2211–2217, 1971.
28. Cowan, R., Whittaker, R. G. Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. Peptide Res. 3:75–80, 1990.
29. Fauchère, J.-L., Pliška, V. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. Eur. J. Med. Chem. 18:369–375, 1983.
30. Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D., Tainer, J. A. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J. Mol. Biol. 228:13–22, 1992.
31. Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. Science 221:709–713, 1983.
32. Connolly, M. L. Depth-buffer algorithms for molecular modelling. J. Mol. Graphics 3:19–24, 1985.
33. Upson, C., Faulhaber, T. Jr., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R., van Dam, A. The application visualization system: A computational environment for scientific visualization. IEEE Comp. Graphics Appl. 9(4):30–42, 1989.
34. Badger, J. Multiple hydration layers in cubic insulin crystals. Biophys. J. 65:1656–1659, 1993.
35. Postorino, P., Tromp, R. H., Ricci, M.-A., Soper, A. K., Neilson, G. W. The interatomic structure of water at supercritical temperatures. Nature (London) 366:668–670, 1993.
36. Stickle, D. F., Presta, L. G., Dill, K. A., Rose, G. D. Hydrogen bonding in globular proteins. J. Mol. Biol. 226: 1143–1159, 1992.
37. Suzuki, S., Green, P. G., Bumgarner, R. E., Dasgupta, S., Goddard III, W. A., Blake, G. A. Benzene forms hydrogen bonds with water. Science 257:942–945, 1992.
38. Hecht, D., Tadesse, L., Walters, L. Correlating hydration shell structure with amino acid hydrophobicity. J. Am. Chem. Soc. 115:3336–3337, 1993.
39. Radzicka, A., Wolfenden, R. Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochemistry 27:1664–1670, 1988.
40. Gibbs, P. R., Radzicka, A., Wolfenden, R. The anomalous hydrophilic character of proline. J. Am. Chem. Soc. 113: 4714–4715, 1991.
41. Eisenberg, D., Weiss, R. M., Terwilliger, T. C., Wilcox, W. Hydrophobic moments and protein structure. Faraday Symp. Chem. Soc. 17:109–120, 1982.
42. Kyte, J., Doolittle, R. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132, 1982.
43. Hagler, A. T., Lifson, S., Dauber, P. Consistent force field studies of intermolecular forces in hydrogen bonded crystals. J. Am. Chem. Soc. 101:5122–5130, 1979.
44. Lesser, G. J., Rose, G. D. Hydrophobicity of amino acid subgroups in proteins. Proteins 8:6–13, 1990.
45. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
46. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., Weng, J. Protein Data Bank. In: "Crystallographic Databases—Information Content, Software Systems, Scientific Applications." Allen, F. H., Bergerhoff, G., Sievers, R. eds. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography, 1987:107–132.
47. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Cryst., Sect. B 39:480, 1983.
48. Thanki, N., Rao, J. K. M., Foundling, S. I., Howe, W. J., Moon, J. B., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Thaisrivongs, S., Wlodawer, A. Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling. Protein Sci. 1:1061–1072, 1992.
49. Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T., Stroud, R. M. Solvent structure in crystals of trypsin determined by x-ray and neutron diffraction. Proteins 12:203–222, 1992.