PROTEINS: Structure, Function, and Bioinformatics 56:102–109 (2004)

Partial Atomic Charges of Amino Acids in Proteins

Annick Thomas,[1*] Alain Milon,[2] and Robert Brasseur[1]
[1]*Centre de Biophysique Moléculaire Numérique (CBMN), Gembloux, Belgium*
[2]*Institut de Pharmacologie et de la Biologie Structurale (CNRS), Toulouse Cedex, France*

ABSTRACT **Using a semiempirical quantum mechanical procedure (FCPAC) we have calculated the partial atomic charges of amino acids from 494 high-resolution protein structures. To analyze the influence of the protein's environment, we considered each residue under two conditions: either as the center of a tripeptide with PDB structure geometry (free) or as the center of 13–16 amino acid clusters extracted from the PDB structure (buried). The partial atomic charges from residues in helices and in sheets were separated. The FCPAC partial atomic charges of the Cβ and Cα of most residues correlate with their helix propensity, positively for Cβ and negatively for Cα (r^2 = 0.76 and 0.6, respectively). The main consequence of burying residues in proteins is the polarization of the backbone C=O bond, which is more pronounced in helices than in sheets. The average shift of the oxygen partial charges that results from burying is −0.120 in helix and −0.084 in sheet with the charge of the proton as unit. Linear correlations are found between the average NMR chemical shifts and the average FCPAC partial charges of Cα (r^2 = 0.8–0.85), N (r^3 = 0.67–0.72), and Cβ (r^2 = 0.62) atoms. Correlations for helix and β-sheet FCPAC partial charges show parallel regressions, suggesting that the charge variations due to burying in proteins differentiate between the dihedral angle effects and the polarization of backbone atoms. Proteins 2004;56:102–109.**
© 2004 Wiley-Liss, Inc.

Key words: partial atomic charges; amino acids; proteins; H-bond; helix; sheet

INTRODUCTION

It is clear that significant progress in the biotechnology of proteins is expected in the coming years. Proteomics would largely benefit from knowing how proteins fold from sequences. Acquiring this knowledge remains a major challenge. Our understanding of the process is restricted by a series of points, among which is an unsatisfactory ability to simulate electrostatic forces. As they fold, proteins swap solvent interactions for intramolecular interactions, which often involve bond polarization such as the H-bonds. The main chain H-bonds implicate 64–70% of the amino acids of sequences to make up the noncovalent architecture of helices and sheets.[1,2] Both structures are elongated because of H-bonds ropes that are almost perpendicular to the sequence chain, along the helix axis and across the sheet strands (Figs. 1 and 2). The basic motif of an H-bond rope encompasses two residues from a first H-bond, the backbone polar O atom, the C=O double bond, the C–N peptide bond, and the N and H atoms to the next H-bond. It is because of the planar geometry of the peptide bond and of the π electrons of the C=O bond that the rope is polarizable. This is especially true in helices because each helix has three parallel H-bond ropes running in the same direction (Fig. 1). In the β-sheet, the H-bond ropes are also parallel, but every second rope runs in the opposite direction, thus impairing structure polarization (Fig. 2). Therefore, bond polarization and partial atomic charges should be different in helices and sheets. The differences are used to explain why N-ends of helices are preferential sites for phosphorylation. The differences might also explain why the NMR chemical shifts differ with secondary structures.[3,4] They could be implicated (as either cause or consequence) in the different geometry of the H-bond node in helices and sheets.[2,5,6] In the β-sheet, the NH · · · O bond has a canonical H-bond geometry because the hydrogen attacks toward the oxygen's sp^2 orbital. In the helix, the hydrogen attacks about 60° above the sp^2 plane, suggesting that the NH · · · O interaction could be modified by the C=O bond polarization.

Bond polarization is not mimicked in most molecular modeling trials. Force fields, such as CHARMM[7] or AMBER,[8] which are used for the refinement of protein structures, have constant partial atomic charges irrespective of secondary structures. Recent progress in methods and in computer facilities have led several groups to develop fast semiempirical quantum mechanical procedures for calculating partial atomic charges of proteins.[9–10] One of these procedures was used to calculate the static and dynamic structures of crambin.[9] The results suggested significant fluctuations of partial charges of main-chain atoms, especially in helices. We used our algorithm to mimic the C=O

The Supplementary Materials Referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html

Grant sponsor: Interuniversity Poles of Attraction Programme, Grant Sponsor: Belgian State, Prime Minister's Office; Grant Sponsor: Federal Office for Scientific, Technical, and Cultural Affairs; Grant number: P5/33; Grant Sponsor: Ministère de la Région Wallonne; Grant numbers: 14540 (PROTMEM) and 0215223 (NANOSENS).

*Correspondence to: Annick Thomas, Centre de Biophysique Moleculaire Numerique (CBMN), Passage des déportés, 2, B-5030 Gembloux, Belgium. E-mail: thomas.a@fsagx.ac.be

Received 1 July 2003; Accepted 5 December 2003

Published online 16 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20093

© 2004 WILEY-LISS, INC.

bond polarization induced when a linear series of formamide spatially arranged as an H-bond rope.[10]

For this article, we calculated the partial atomic charges of residues from a large series of high-resolution protein structures.[11] The aim was not to test how charges varied during molecular dynamics but to analyze charge diversity, essentially with respect to residues and secondary structures. We compared the partial atomic charges when each amino acid was the center of a tripeptide with its two sequence neighbours (free) and when it was the centre of a large block of residues extracted from the PDB structures (buried). In both sets, the residue geometry was the same as in the three-dimensional (3D) structures only the environment was different.

There is no direct way of assessing calculated partial charges by experimentation. The partial charges in force fields are validated by comparing the thermodynamic parameters extracted from molecular dynamic simulations to experimental measurements. However, in the derivation of the ECEPP force field,[12] the partial charges on the Cα and Cβ atoms were claimed to correlate with the NMR chemical shift, suggesting that correlation with the NMR chemical shifts could serve as an experimental validation. Recently, improved accuracy and a greater number of NMR structures have led to the creation of reliable databases of chemical shifts in secondary structures.[4] We took advantage of this development to compare the partial atomic charges calculated with FCPAC with the shifts in these databases.

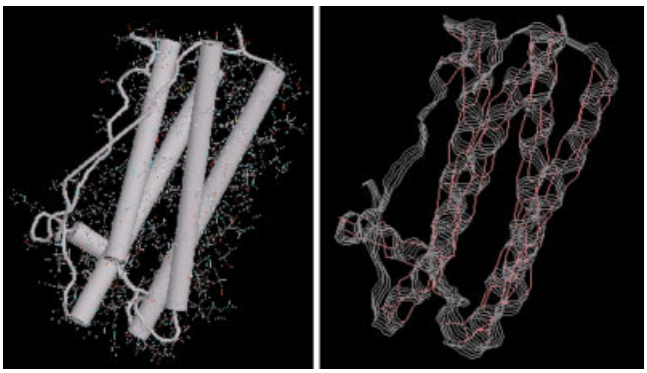

Fig. 1. PDB structure of 1BGCH. The protein is a helix bundle. **Left:** Diagramatic view of the 3D structure. **Right:** The H-bond network (ropes in pink) was added overtop of the skeleton (gray ribbon). The network starts when two amino acids are involved in a main-chain H-bond: the rope is made up of the main-chain H-bond plus the O=C of a residue, the peptide bond, and the NH of its nearest partner in the sequence.

MATERIALS AND METHODS

Protein Structure Data Set

We used a data set of 494 non homologous structures with a resolution better than 1.8 Å.[11] Although hydrogen atoms cannot be observed directly in most of these crystal structures, Word et al.[13,14] showed that projected hydrogen atoms from high-resolution structures are physically meaningful.

Secondary structures (α-helix and β-sheet cores) were defined in terms of both ϕ-ψ angles and H-bond patterns as previously described.[2] Residues in helix core are H-bonded with their n−4 and n+4 sequence neighbors. Residues in sheet are H-bonded with strand partners on both their NH and C=O moieties. Of course, the proline residues are H-bonded only on their CO side. Partial charges were classified according to types of residues and secondary structures (see Supplementary Materials).

Partial Charges

The charges of each residue have been evaluated twice. In the first series (free residues in Table I and Supplementary Table II), the environment of each atom was a combination of all the amino acid atoms plus all the atoms of the −1 and +1 neighbors in the sequence. In the second series (buried in Supplementary Tables I and II), the environment was all atoms from the amino acid plus all the atoms from any amino acid <10 Å away in the 3D structure. In helices, this corresponded to blocks of 13 ± 1 amino acids for alanine and of 16 ± 2 for tryptophan. All atomic charges were collected in Pex files, together with secondary structures, phi/psi and a series of parameters such as the solvent-accessible surface.[15,16] Data in Tables 1 and 2 account for the mean values ± SD of partial atomic

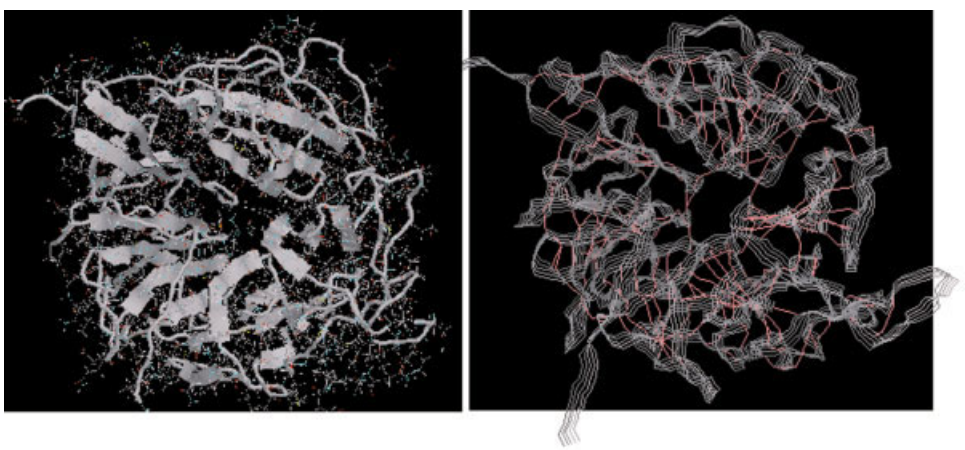

Fig. 2. PDB structure of 1A12AH. The protein is a β-barrel; the legend is the same as for Figure 1.

TABLE I. Partial Atomic Charges of the N, HN, Cα, C′, O, and Cβ Atoms of Amino Acids[†]

N	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	−0.416	−0.621	0.024	−0.592	0.033	−0.598	0.027	−0.588	0.034
ARG	−0.348	−0.627	0.024	−0.590	0.033	−0.595	0.026	−0.586	0.034
ASN	−0.416	−0.622	0.026	−0.585	0.037	−0.589	0.026	−0.576	0.036
ASP	−0.516	−0.599	0.027	−0.568	0.040	−0.574	0.027	−0.560	0.044
CYS	−0.416	−0.600	0.025	−0.558	0.041	−0.577	0.029	−0.550	0.038
GLN	−0.416	−0.617	0.025	−0.584	0.034	−0.587	0.028	−0.578	0.038
GLU	−0.416	−0.609	0.024	−0.575	0.036	−0.584	0.028	−0.564	0.038
GLY	−0.416	−0.594	0.025	−0.566	0.036	−0.575	0.026	−0.563	0.034
HIS	−0.416	−0.615	0.023	−0.584	0.036	−0.589	0.031	−0.577	0.040
ILE	−0.416	−0.616	0.020	−0.590	0.037	−0.601	0.027	−0.587	0.035
LEU	−0.416	−0.617	0.022	−0.593	0.034	−0.600	0.026	−0.586	0.034
LYS	−0.416	−0.624	0.027	−0.585	0.034	−0.592	0.023	−0.586	0.034
MET	−0.416	−0.616	0.024	−0.592	0.032	−0.602	0.025	−0.583	0.037
PHE	−0.416	−0.616	0.026	−0.589	0.038	−0.603	0.027	−0.581	0.035
PRO	−0.255	−0.347	0.027	−0.310	0.030	−0.343	0.021		
SER	−0.416	−0.617	0.025	−0.579	0.039	−0.595	0.026	−0.571	0.037
THR	−0.416	−0.606	0.023	−0.573	0.037	−0.589	0.025	−0.569	0.036
TRP	−0.416	−0.612	0.024	−0.589	0.034	−0.602	0.025	−0.575	0.034
TYR	−0.416	−0.612	0.025	−0.586	0.035	−0.599	0.026	−0.578	0.035
VAL	−0.416	−0.614	0.020	−0.588	0.036	−0.602	0.026	−0.587	0.034
HN	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	0.272	0.307	0.025	0.340	0.042	0.358	0.027	0.352	0.048
ARG	0.275	0.306	0.027	0.330	0.046	0.348	0.034	0.334	0.056
ASN	0.272	0.308	0.031	0.335	0.048	0.355	0.039	0.327	0.064
ASP	0.294	0.310	0.035	0.331	0.046	0.355	0.028	0.336	0.043
CYS	0.272	0.306	0.021	0.335	0.047	0.355	0.025	0.315	0.063
GLN	0.272	0.304	0.034	0.336	0.044	0.352	0.033	0.337	0.051
GLU	0.272	0.306	0.036	0.338	0.041	0.356	0.026	0.339	0.047
GLY	0.272	0.304	0.025	0.335	0.046	0.363	0.030	0.346	0.038
HIS	0.272	0.306	0.027	0.333	0.047	0.353	0.029	0.339	0.053
ILE	0.275	0.305	0.031	0.325	0.050	0.341	0.038	0.327	0.061
LEU	0.272	0.305	0.030	0.329	0.048	0.349	0.032	0.328	0.063
LYS	0.272	0.299	0.031	0.332	0.044	0.353	0.028	0.340	0.051
MET	0.272	0.299	0.047	0.332	0.047	0.354	0.029	0.331	0.060
PHE	0.272	0.301	0.036	0.328	0.048	0.348	0.031	0.331	0.054
SER	0.272	0.306	0.027	0.332	0.045	0.350	0.027	0.342	0.048
THR	0.272	0.307	0.027	0.329	0.046	0.344	0.027	0.336	0.058
TRP	0.272	0.307	0.026	0.326	0.051	0.350	0.041	0.324	0.058
TYR	0.272	0.307	0.024	0.326	0.051	0.350	0.028	0.325	0.063
VAL	0.272	0.306	0.030	0.328	0.049	0.346	0.031	0.331	0.062
CA	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	0.034	0.267	0.019	0.267	0.013	0.272	0.011	0.264	0.012
ARG	−0.264	0.247	0.018	0.240	0.013	0.243	0.010	0.238	0.010
ASN	0.014	0.250	0.024	0.244	0.016	0.243	0.014	0.239	0.015
ASP	0.038	0.253	0.022	0.246	0.016	0.246	0.010	0.242	0.016
CYS	0.021	0.191	0.017	0.186	0.014	0.188	0.009	0.183	0.016
GLN	−0.003	0.238	0.025	0.234	0.013	0.236	0.012	0.233	0.013
GLU	0.040	0.236	0.021	0.234	0.013	0.236	0.009	0.228	0.014
GLY	−0.025	0.115	0.021	0.108	0.018	0.119	0.017	0.103	0.015
HIS	−0.058	0.240	0.022	0.235	0.013	0.235	0.014	0.234	0.013
ILE	0.060	0.218	0.023	0.214	0.014	0.217	0.013	0.213	0.013
LEU	−0.052	0.243	0.023	0.242	0.014	0.245	0.012	0.239	0.014
LYS	−0.240	0.243	0.025	0.238	0.014	0.242	0.011	0.237	0.014
MET	−0.024	0.244	0.026	0.244	0.014	0.249	0.011	0.240	0.016
PHE	−0.002	0.234	0.024	0.235	0.015	0.239	0.014	0.234	0.014
PRO	−0.027	0.169	0.020	0.170	0.014	0.166	0.009		
SER	−0.025	0.208	0.020	0.193	0.014	0.197	0.013	0.190	0.014
THR	−0.039	0.187	0.019	0.169	0.014	0.173	0.012	0.168	0.014
TRP	−0.028	0.238	0.022	0.233	0.017	0.234	0.020	0.229	0.018
TYR	−0.001	0.237	0.018	0.235	0.015	0.238	0.012	0.233	0.015
VAL	−0.088	0.220	0.022	0.215	0.014	0.220	0.012	0.214	0.013
C′	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	0.597	0.582	0.031	0.560	0.025	0.559	0.020	0.562	0.023
ARG	0.734	0.582	0.030	0.558	0.026	0.560	0.023	0.557	0.024
ASN	0.597	0.585	0.031	0.561	0.029	0.559	0.019	0.564	0.023

TABLE I. (Continued)

ASP	0.537	0.598	0.032	0.570	0.027	0.573	0.024	0.570	0.026
CYS	0.597	0.588	0.035	0.574	0.030	0.571	0.028	0.584	0.023
GLN	0.597	0.586	0.029	0.562	0.024	0.565	0.020	0.563	0.022
GLU	0.537	0.598	0.026	0.568	0.025	0.569	0.022	0.569	0.023
GLY	0.597	0.600	0.028	0.581	0.028	0.575	0.025	0.580	0.023
HIS	0.597	0.590	0.028	0.563	0.026	0.571	0.023	0.564	0.022
ILE	0.597	0.588	0.028	0.560	0.027	0.562	0.025	0.565	0.024
LEU	0.597	0.585	0.027	0.557	0.027	0.558	0.023	0.563	0.023
LYS	0.734	0.579	0.035	0.561	0.025	0.560	0.021	0.562	0.021
MET	0.597	0.575	0.041	0.559	0.027	0.559	0.025	0.563	0.025
PHE	0.597	0.592	0.025	0.559	0.030	0.569	0.025	0.559	0.026
PRO	0.590	0.594	0.025	0.567	0.035	0.555	0.022		
SER	0.597	0.590	0.029	0.567	0.026	0.562	0.021	0.566	0.022
THR	0.597	0.587	0.031	0.565	0.026	0.563	0.021	0.568	0.026
TRP	0.597	0.586	0.036	0.565	0.029	0.568	0.027	0.562	0.025
TYR	0.597	0.591	0.029	0.561	0.028	0.573	0.024	0.561	0.024
VAL	0.597	0.587	0.026	0.560	0.026	0.559	0.024	0.563	0.024
O	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	−0.568	−0.574	0.042	−0.673	0.069	−0.701	0.049	−0.661	0.058
ARG	−0.589	−0.570	0.042	−0.658	0.071	−0.682	0.051	−0.658	0.065
ASN	−0.568	−0.581	0.041	−0.664	0.075	−0.693	0.048	−0.656	0.064
ASP	−0.582	−0.605	0.042	−0.673	0.077	−0.698	0.043	−0.665	0.073
CYS	−0.568	−0.585	0.039	−0.694	0.063	−0.718	0.051	−0.695	0.049
GLN	−0.568	−0.579	0.043	−0.666	0.068	−0.687	0.048	−0.673	0.060
GLU	−0.582	−0.593	0.043	−0.681	0.069	−0.705	0.050	−0.676	0.063
GLY	−0.568	−0.565	0.041	−0.668	0.076	−0.706	0.059	−0.653	0.055
HIS	−0.568	−0.586	0.043	−0.663	0.075	−0.690	0.059	−0.657	0.067
ILE	−0.568	−0.574	0.040	−0.657	0.076	−0.698	0.054	−0.655	0.062
LEU	−0.568	−0.574	0.039	−0.663	0.079	−0.699	0.051	−0.649	0.065
LYS	−0.589	−0.570	0.046	−0.652	0.069	−0.683	0.050	−0.662	0.056
MET	−0.568	−0.575	0.037	−0.665	0.075	−0.706	0.050	−0.654	0.063
PHE	−0.568	−0.572	0.042	−0.656	0.075	−0.688	0.052	−0.644	0.071
PRO	−0.548	−0.580	0.044	−0.670	0.077	−0.727	0.046		
SER	−0.568	−0.578	0.039	−0.671	0.069	−0.698	0.049	−0.668	0.057
THR	−0.568	−0.574	0.039	−0.659	0.072	−0.690	0.049	−0.663	0.063
TRP	−0.568	−0.585	0.044	−0.669	0.072	−0.697	0.060	−0.661	0.074
TYR	−0.568	−0.568	0.041	−0.662	0.073	−0.690	0.048	−0.663	0.063
VAL	−0.568	−0.568	0.036	−0.661	0.072	−0.695	0.051	−0.657	0.064
CB	Amber95	FREE	SD	BURIED	SD	HELIX	SD	SHEET	SD
ALA	−0.183	−0.416	0.010	−0.392	0.017	−0.400	0.013	−0.380	0.013
ARG	−0.001	−0.224	0.010	−0.215	0.016	−0.224	0.013	−0.206	0.014
ASN	−0.204	−0.249	0.014	−0.238	0.018	−0.247	0.013	−0.233	0.015
ASP	−0.030	−0.299	0.015	−0.276	0.017	−0.285	0.013	−0.271	0.016
CYS	−0.123	0.037	0.013	0.096	0.022	0.086	0.022	0.103	0.020
GLN	−0.004	−0.229	0.012	−0.212	0.017	−0.221	0.014	−0.203	0.013
GLU	0.056	−0.230	0.014	−0.207	0.017	−0.214	0.015	−0.199	0.013
HIS	−0.007	−0.160	0.012	−0.153	0.022	−0.167	0.020	−0.145	0.019
ILE	0.130	−0.042	0.011	−0.032	0.011	−0.039	0.008	−0.027	0.009
LEU	−0.110	−0.251	0.015	−0.235	0.017	−0.245	0.014	−0.227	0.016
LYS	−0.009	−0.222	0.011	−0.214	0.016	−0.224	0.012	−0.204	0.013
MET	0.003	−0.290	0.012	−0.274	0.015	−0.283	0.012	−0.266	0.014
PHE	−0.034	−0.168	0.014	−0.156	0.020	−0.171	0.015	−0.148	0.017
PRO	−0.007	−0.209	0.008	−0.196	0.011	−0.207	0.010		
SER	0.212	0.075	0.012	0.079	0.017	0.071	0.014	0.085	0.015
THR	0.365	0.225	0.011	0.232	0.012	0.225	0.009	0.238	0.010
TRP	−0.005	−0.146	0.011	−0.133	0.021	−0.142	0.020	−0.124	0.017
TYR	−0.015	−0.167	0.011	−0.148	0.020	−0.165	0.015	−0.138	0.016
VAL	0.299	−0.016	0.011	−0.005	0.012	−0.013	0.008	0.000	0.009

[†]The atoms are listed by their PDB codes. The partial charges of Amber are taken from W. D. Cornell et al.[8] (Current parameter values are available from http://www.amber.ucsf.edu/amber/amber.html). The partial charges were calculated from 494 PDB structures using FCPAC.[10] Two sets of residue environments were considered. In the free set, the residue had the PDB geometry but only two neighbors its −1 and its +1 partners in the structure. In the buried set, the amino acids were extracted from the PDB structure within blocks of 13–16 residue partners as explained in Materials and Methods. All the atoms of the block were taken into account for the calculation, but the partial atomic charge of the central residue was the only one reported in the average calculation: values are means ± SD. From the buried set, we separated the residues in helix (H-bond with a n−4 and n+4 neighbors) from the residues in sheets.

charges from the residues that are <30% accessible to the solvent in a 3D structure; 99% of these residues had backbone atoms that were inaccessible to water.

FCPAC (Fast Calculation of Partial Atomic Charges)

The method was described previously.[10] To each atom in an organic molecule, a chemist allocates a set of properly hybridized orbitals and connects some of them to make chemical bonds. A polarity can be associated with each bond according to the difference in electronegativity between the bonded atoms. For systems including only σ bonds and lone pairs, these empirical methods can give reliable partial atomic charges. But when one tries to deal with conjugated π systems, calculations refer to quantum chemistry. Diverse approaches have been developed, including the linear combination of atomic orbitals (LCAO). This approach, initially developed by Hückel,[17] has allowed the qualitative treatment of delocalized π system. A more elaborate method called PPP[18,19] (Pariser–Parr–Pople) is currently used to treat π electrons in force fields. The success of the method is mainly due to the hypothesis that σ and π electrons can be treated independently. This concept is explained as follows: if orbitals have slight overlaps, the exchange of energy is low and can be disregarded at first. This criterion holds for σ and π systems but is also satisfied for localized molecular orbitals. In this case, the overlap is slight even between bond-bond and bond- lone pair orbitals. We use these hypotheses to create FCPAC where all σ electrons are included in the calculation (in contrast to PPP). We first create a set of hybrid atomic orbitals and connect them to make single bonds and conjugated systems. Hybridization levels of orbitals are calculated by using a quantitative expression of Bent's rule as in the VALBOND method. Each bond, lone pair, and π system is assumed to be a block that overlaps very slightly with other blocks. Finally, a LCAO process is applied to each block taking into account the mean field of the charges from the other blocks. The process is iterated to the point of convergence. The calculations were performed at complete neglect of differential overlap (CNDO) level. The method, initially parametrized on small organic compounds to reproduce the HF/6-31G(d,p) Mulliken charges,[10] was extended to fit electrostatic potential (Merz-Kollman-Singh) resulting from RB3LYP/6- 311++(2d,2p) electronic density. Ab initio calculations were performed by using GAUSSIAN98.

RESULTS

In proteins, the solvent accessibility of residues varies: 71% of Ala, 87% of Leu, 87% of Ile, and 85% of Val have surfaces that are 70% inaccessible to water. On the other hand, only 30% of Lys, 34% of Glu, and 48% of Arg residues have the same solvent inaccessibility. These data concern the whole residues, but when we limit them to backbones, all residues except Gly and Pro have a similar pattern, meaning that 92 ± 3% of the residue backbone have at least 70% of their surface inaccessible to water. The percentages for Gly and Pro are 67% and 69%, respectively. For this article, we analyzed the residues whose total surface was <30% accessible to water. For these residues, the backbone is completely buried. We calculated the charges for 4832 Ala, 2207 Arg, 1082 Asn, 904 Asp, 212 Cys, 1630 Gln, 1000 Glu, 1833 Gly, 745 His, 3185 Ile, 5046 Leu, 2692 Lys, 1130 Met, 2047 Phe, 2937 Pro, 1976 Ser, 3117 Thr, 595 Trp, 1964 Tyr, and 4400 Val.

Partial Atomic Charges of Free Residues

In the absence of a proteic environment, the FCPAC partial charges of backbone atoms are similar for all residues (except the N of proline −0.347) (Table I). Mean values are 0.305 ± 0.030 for HN, −0.613 ± 0.024 for N, 0.224 ± 0.022 for Cα, −0.578 ± 0.041 for O, and 0.588 ± 0.030 for C, respectively. For reference, note here that in AMBER, the HN and N charges are 0.272 and −0.416 except for Asp, Arg, and Pro. The average Cα value is −0.033, slightly negative, whereas it is positive in FCPAC. The C=O bond charges are similar in AMBER and FCPAC except for Asp, Glu, Pro, Arg, and Lys, the atomic charges of which do not depart from average values in FCPAC as they do in AMBER.

The partial atomic charges of side-chain atoms are specific to each residue. For instance, the charges of the Cα and Cβ vary with side chains (See Table I and Supplementary Table II).

Correlation With Secondary Structures

The FCPAC partial charges of most residues Cα and Cβ are linearly correlated with their helix propensity in proteins (alpha helix propensities) (Fig. 3). The regression coefficient is 0.60 for Cα and 0.76 for Cβ when the Asp and Asn residues are omitted. This was accepted because of the shortness and polarity of these residue side-chains. No such correlation is seen for the AMBER Cα and Cβ atomic charges (Fig. 3).

It is of interest that the regression is positive for Cα and negative for Cβ, suggesting that a more positive charge of Cα and a more negative charge of Cβ favor an alpha helix structure.

Partial Atomic Charges of Amino Acids in Proteins

The embedding of residues in proteins results in small changes of charge values, but this includes different situations when side-chain or main-chain atoms are considered (Table I and Supplementary Table II; compare buried with free).

There is essentially no change in the partial charges of side-chain atoms, but some change in the partial charges of the main-chain atoms. With burying, the Cβ and Cα partial charges vary by 0.016 and 0.005, respectively. Changes are 0.026, 0.031, and 0.024 for HN, N, and C′, respectively, and reach 0.088 for the main-chain oxygen (Table I and Fig. 4). Hence, oxygen is the most polarized atom. This global shift further includes differences between secondary structures (Fig. 4 and compare "Buried" with "Helix" and "Sheet" in Table I). The shift of the main-chain oxygen partial charges is greater in the helix than in the sheet. The mean shift is −0.120 in the helix

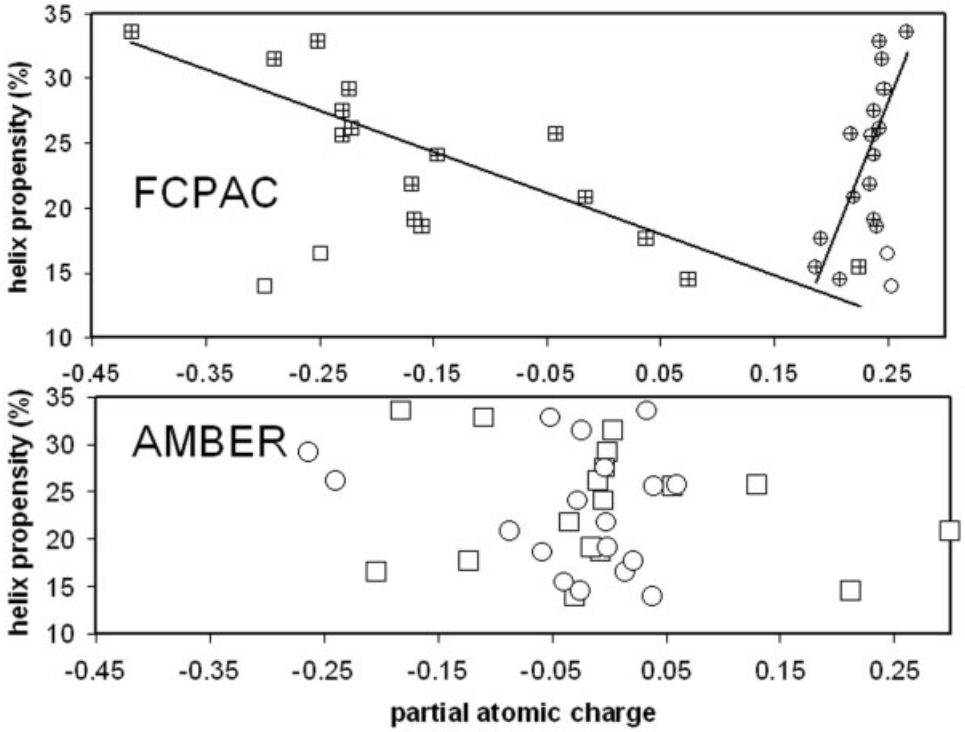

Fig. 3. Relationships between the average helix propensity of amino acids and their average Cβ and Cα partial charges. The helix frequency was calculated from the database as previously reported.[15] The top plot is of FCPAC average partial charges, and the bottom plot of AMBER partial charges. The square symbols represent the Cβ partial charge and the circles the Cα partial charges. The symbols containing a cross were used for calculating the linear regressions. It includes Ala, Arg, Cys, Gln, Glu, Ile Leu, Lys, Met, Ser, Thr, Val, Trp, Phe, Tyr, and His. For Cb, the regression coefficient is $r^2 = 0.76$, and for Cα $r^2 = 0.60$.

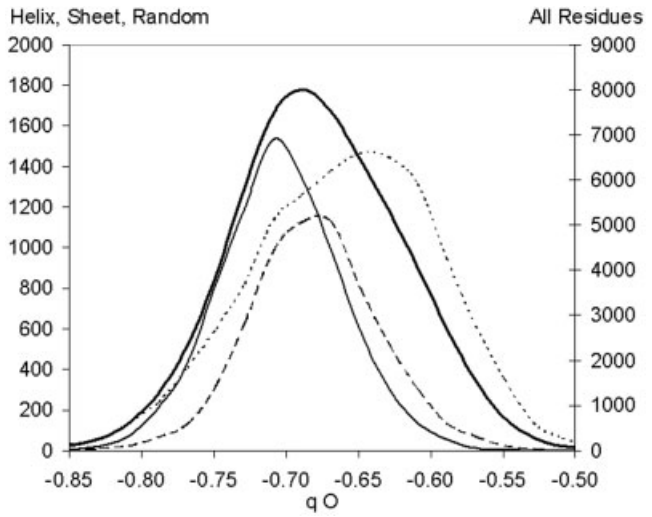

Fig. 4. Distribution of the partial oxygen charges in different secondary structures. The plots show the distribution frequency (in %) of the partial main-chain oxygen charges of all "buried" residues (58509 values, thicker line, scale on the right y axis) but also those of residues in sheet (6274 values ----), α helix (7706 values ____) and in random coil (12437 values . . ..) structures (scale on the left y axis).

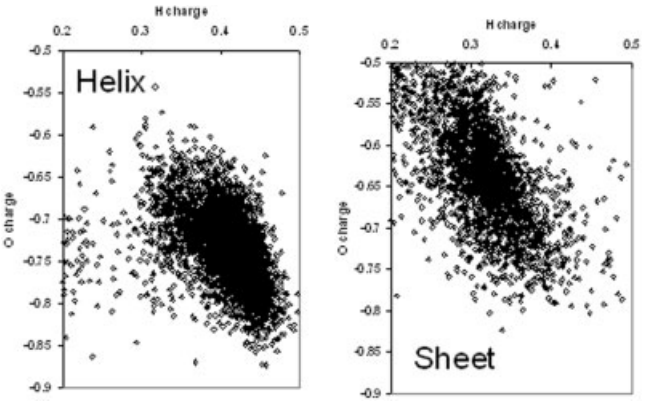

Fig. 5. Correlation between FCPAC partial charges of oxygen and HN in helices and sheets. The oxygen is from alanine and the HN are from the residues H-bonded on the C=O side of the alanine.

and −0.084 in the sheet. This is accompanied by greater variation of the HN partial charge (0.046 in helix) (Fig. 5).

A rapid calculation of the Coulomb energy point charge suggests that the polarization of the O and HN could result in a δE fluctuation of -0.49 ± 0.36 Kcal per bond in helices and -0.23 ± 0.42 kcal per bond in sheets with respect to the free charges (ε_{cst} = distance).

Correlation With NMR Chemical Shifts

We used the NMR chemical shift values of amino acid atoms taken from the database published by Wang and Jardetzky.[4] The charges for helix and sheet residues correlated with their respective average NMR chemical

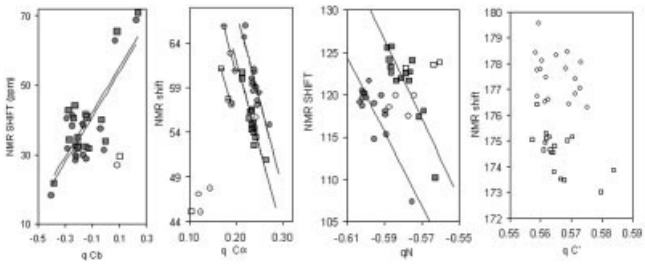

Fig. 6. Relationships between the average chemical shifts of atoms and their average FCPAC partial charges. The chemical shifts are taken from Wang and Jardetsky.[4] The FCPAC partial charges are from helix and sheet columns of Table I. The plots are, from left to right, for Cβ, Cα, N, and C′. The square symbols are for sheets and circles for helices. The gray shaded symbols were used for linear regressions. Analysis of the regressions: for Cβ, Cys was omitted and regression coefficients (r^2) of 0.62 were observed for helix and sheet; for Cα, Gly was omitted; Ser, Thr, and Cys (symbols with a — inside) were correlated with a r^2 of 1 for helix and 0.96 for sheet; the other residues were correlated with a regression coefficient of $r^2 = 0.80$ for helix and $r^2 = 0.85$ for sheet; for N, Asp, Glu, Gln, and Cys were excluded and the linear regressions had $r^2 = 0.67$ for helix and $r^2 = 0.72$ for sheet. For C′, the linear regressions had an r^2 coefficient of <0.5 and were not considered.

shifts. Good correlations were found between the Cα chemical shifts and their FCPAC partial charges assuming that glycine is not included in the correlation and that Ser, Thr, and Cys are correlated separately with a regression coefficient of $r^2 = 0.96–1$ (Fig. 6). The main correlation gathers all the other residues (Ala, Arg, Asn, Asp, Glu, Gln, His Ile, Leu, Lys Met, Phe, Trp, Tyr, Val) and gives parallel linear correlations for the helix and the sheet with regression coefficients >0.8. The linear regressions for N and Cβ chemical shifts versus FCPAC charges in helices and in sheets are also parallel (Fig. 6). No significant correlation was observed for C′ and HN (Fig. 6). Carbon and nitrogen NMR chemical shifts are influenced both by dihedral angle effects (local conformations) and by bond polarization. The plots seem to clearly separate both effects, each regression line illustrating the dihedral angle effects and the distance between parallel regressions being attributed to the bond polarization. Although correlated, the regression is poorer for Cβ than for Cα. This might arise from the dispersion of dihedral angles around Cβ, due to the heterogeneity in side-chain conformations. Here, helix and sheet clustering does not help, and the clustering of each amino acid and of each χi dihedral angle could be more successful. We have no explanation for the lack of significant correlation between the partial charges of C′ and their chemical shifts. No significant correlation is observed for either Cβ, Cα, C′, N, or HN when the AMBER partial charges are tested.

The correlations observed with the Cα and, to a lesser extent, with the Cβ and N are strong indications that the FCPAC-derived partial charges are approaching the real polarization for protein cores.

CONCLUSIONS

Recently, thanks to the improvements in computer capacities, semiempirical quantum procedures for fast calculation of charges have been developed. Herein we analyzed the changes in partial atomic charges of amino acids that occur with their burying in different secondary structures of a protein environment.

One conclusion is that, in the absence of any environment, the partial charges of the main-chain atoms of all residues are similar, whereas the partial charges of side-chain atoms (from Cα and over) vary. We have not analyzed whether they varied with χi dihedral angles. From there on and for the sake of clarity, we analyzed the results from two points of view: what is intrinsic to residues and what appears with protein folding. Intrinsic to residues is the correlation between their partial charges of Cα and Cβ and their helix propensities. A previous series of articles showed that in sheets, the CαH approaches the main-chain oxygen acting as the NH counterpartner in the pitching of backbone oxygen.[2,5,6,20] We demonstrated that in helices, the CβH is often the CαH substitute in that respect.[2] Hence, finding that residues with more negative charged Cβ and more positively charged Cα have a higher tendency to be present in helices is interesting, because it suggests an electrostatic CH$\cdots$O interaction.

Looking at the consequences of residue burying in a protein environment, most changes concern the backbone atoms and especially, the π electrons of the C=O bonds. These effects are different when they occur in sheets than in helices which, as predicted, tolerate a higher polarization because of their three parallel and unidirectional H-bond ropes per helix. The charge fluctuations of Cα, Cβ, and N are correlated with the NMR chemical shifts, and this is an important point suggesting that FCPAC values account for part of the polarization problem. However, we have no explanation as to why C′ charges do not correlate with the chemical shifts.

ACKNOWLEDGMENTS

A. Thomas is directeur de recherches INSERM (France), and R. Brasseur is directeur de recherches FNRS (Belgium). The authors acknowledge B. K. Ho and M. Swinnen for their participation in the manuscript.

REFERENCES

1. Stickle DF, Presta LG, Dill KA, Rose G. Hydrogen bonding in globular proteins, J Mol Bio 1992;226:1143–1159.
2. Thomas A, Benhabiles N, Meurisse R, Ngwabije R, Brasseur R. Pex, an analytical tool for PDB files. II H-Pex: non-canonical H-bonds in alpha helices. Proteins 2001;43:37–44.
3. Case DA. Interpretation of chemical shifts and coupling constants in macromolecules. Curr Opin Struct Biol 2000;10:197–203.
4. Wang Y, Jardetzky O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 2002;11:852–861.
5. Derewenda ZS, Lee L, Derewend; U. The occurrence of C–H–O H-bonds in proteins. J Mol Biol 1995;252:248–262.
6. Ho BK, Curmi PMG. Twist and shear in β-sheets and β-ribbons. J Mol Biol 2002;317:291–308.
7. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau, FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yi D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem 1998;B102:3586–3616.

8. Cornell WD, Cieplak P, Bayly CL, Gould IR, Mer KM Jr, Ferguson D, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 1995;117:5179–5197.
9. Liu H, Elstner M, Efthimios K, Frauenheim T, Hermans J, Yang W. Quantum mechanics simulation of protein dynamics on long timescale. Proteins 2001;44;484–489.
10. Swinnen M, Thomas A, Brasseur R. FCPAC: fast calculation of partial atomic charges using strictly localised molecular orbitals. Comput Mater Sci 2002;25:4, 590–595.
11. Lovell SC, Davis IW, Bryan Arendall W III, de Bakker PW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 2003;50:437–450.
12. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 1975;79:2361–2381.
13. Word MJ, Lovell SG, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol 1999;285:1711–1733.
14. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of sidchain amide orientation. J Mol Biol 1999;285:1735–1747.
15. Thomas A, Bouffioux O, Geeurickx D, Brasseur R. PEX, analytical tools for PDB files. I. Pex, analytical tools for PDB files. I. GF-Pex: the basic file to describe a protein. Proteins 2001;43:28–36.
16. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. Protein Sci 2003. Forthcoming.
17. Hückel E. Quanten theorische Beiträge zum Benzolproblem. I. Die electron enkonfiguration des Benzols. Z Phys 1931;70:204–286.
18. Pariser R, Parr RG. A semi empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I. J Chem Phys 1953;21:466-471.
19. Pariser R, Parr RG. A semi empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II. J Chem Phys 1953;21:767-776.
20. Desiraju GR. The C–H–O hydrogen bond in crystals: what is it? Acc Chem Res 1991;24:290–296.
21. Halgren TA. The Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 1996;17:490–519.