Journal of Computer-Aided Molecular Design, 5 (1991) 545–552
ESCOM

J-CAMD 141

HINT: A new method of empirical hydrophobic field calculation for CoMFA

Glen E. Kellogg*, Simon F. Semus and Donald J. Abraham

Department of Medicinal Chemistry and Division of Biomedical Engineering, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0540, U.S.A.

Received 8 December 1990
Accepted 2 July 1991

Key words: Comparative Molecular Field Analysis; HINT; Hydrophobicity; Hydrophobic atom constant; PLS

SUMMARY

An empirical hydrophobic field-like 3D function has been calculated with the program HINT (hydrophobic interactions) and imported into the SYBYL implementation of CoMFA (Comparative Molecular Field Analysis). The addition of hydrophobicity appears to offer increased chemical interpretability of CoMFA models. An example is given using the steroid model reported by Cramer et al. (J. Am. Chem. Soc., 110 (1988) 5959). While addition of the HINT field did not improve statistical parameters in this model, the CoMFA coefficient contours from the hydrophobic field unambiguously define the most active steroid molecules in the chemical terms of hydrophobic and polar substituents.

INTRODUCTION

In the short time since the Comparative Molecular Field Analysis (CoMFA) was initially described by Cramer et al. [1], the method has become a popular and valuable tool for computer-aided drug design [2–5]. The standard CoMFA method samples steric and electrostatic fields surrounding a set of drugs/ligands to determine a quantitative structure-activity relationship (QSAR). To address the importance of hydrophobicity in QSAR, we have integrated an empirical hydrophobic field into CoMFA and show that for drug-design purposes the three-field analysis may have improved predictive and interpretive capabilities over the standard two-field analysis. Steroid binding to corticosteroid-binding globulin is used to illustrate the technique.

Hydrophobicity has long been regarded as a key parameter for understanding drug activity [6,7]. Practically, hydrophobicity is measured and reported as P, the partition coefficient of solubility between water and an organic solvent. Hansch first demonstrated the use of log P and sub-

* To whom correspondence should be addressed.

structure hydrophobic constants as components in QSAR correlation equations in 1964 [8]. Rekker [9] and later Hansch and Leo [10] developed empirical methods for prediction of hydrophobicity by reducing experimental solvent partition data into hydrophobic fragment constants modified by a small collection of situation-dependent factors. Abraham and Leo [11] proposed the use of hydrophobic *atom* constants that encode empirical thermodynamic information to evaluate interactions (i.e., docking) between small and large (biological) molecules, and to evaluate intramolecular interactions such as protein folding. The program HINT (Hydrophobic Interactions) was written to investigate these interactions.

HINT

HINT [12,13] is a program designed to calculate and map the hydrophobic environment of small molecules and proteins. The key parameter is the hydrophobic atom constant, which is the atom-based analog of the fragment constant. HINT calculates hydrophobic atom constants from published hydrophobic fragment constants [10,11] using the following criteria: (1) the sum of atom constants within a fragment equals the fragment constant value; (2) factors [10,11] (e.g., bond, chain-branching, geminal and vicinal halogen, etc.) are applied to all eligible atoms, while polar proximity factors (when necessary) are applied to the central atom of fragments; and (3) terminal (or frontier) atoms are considered by HINT to be more important than central atoms of fragments and are generally maintained at the original atomic values.

The hydrophobic field is calculated by superimposing a grid of test points in space around the molecules of interest and evaluating the net sum at each test point using the following equation

$$A_t = \sum_{i=1}^{drug} s_i a_i R_{it}$$

where s_i and a_i are the solvent-accessible surface area and hydrophobic atom constant for atom i. R_{it} is a function of the distance (r) between atom i and the test point t: $R_{it} = e^{-r}$ for this study [14].

CoMFA

The knowledge of the three-dimensional structure of ligands and receptors is of primary importance for completely understanding the nature of binding interactions. However, in most cases little or no detailed information exists regarding the receptor structure, and this information must be inferred from the ligand's relative molecular shape and activity. The interactions between molecular fragments and a receptor which ultimately lead to binding and biological activity are usually noncovalent in nature. CoMFA [1], by first superimposing a grid of test points around a series of aligned ligands, quantifies these noncovalent interactions in terms of steric and electrostatic fields. Cross-validation in the Partial Least Squares (PLS) analysis [15,16] of the relationship between the steric/electrostatic fields and the biological activity of the ligands maximizes the likelihood that the Field Analysis results have predictive value.

CoMFA is based on the assumptions that structurally diverse compounds bind in conformations that present a common steric, electrostatic and hydropathic shape or pattern to a constant receptor site. These conditions demand strict alignment rules for the ligand set in order to associ-

ate the locality of molecular fragments with changes in activity. The fields employed in CoMFA to determine the interaction between molecular fragments and the receptor are products of molecular mechanics structure parameterization and, hence, provide only steric and electrostatic information. Hydrophobicity, which is not a molecular-mechanics derivable property, is thus not accounted for despite its well-established significance in noncovalent ligand-receptor interactions [17,18]. Quantification of hydrophobic interactions with a hydrophobicity field would overcome this limitation, and provide additional interpretive tools for understanding drug QSAR. Consequently, we have incorporated the HINT hydrophobicity fields into the SYBYL (5.32) implementation of CoMFA [19] to evaluate hydrophobic interactions as a component of 3D QSAR. As an initial example of the technique, this contribution describes a combined HINT/CoMFA study of steroid binding to carrier proteins. This data set, initially examined by Cramer et al. [1] and later by Norinder [5], provides a well-defined overlap model.

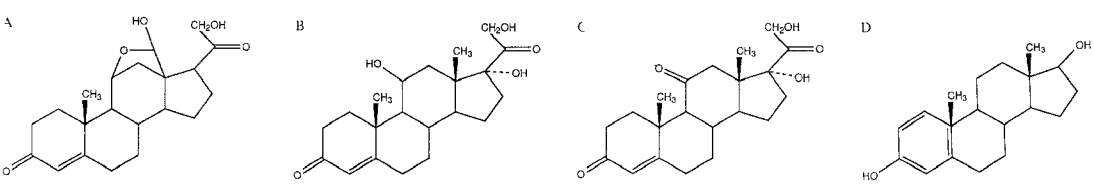
CALCULATIONAL DETAILS

The molecules (21 steroids), their conformations, and calculational conditions employed in this study were taken from the paper of Cramer et al. [1] and are additionally available through the CoMFA tutorial in the SYBYL modeling programs. No additional minimization or alignment procedures were applied to the molecular coordinate data. Hydrophobic maps were calculated with HINT using all atoms in the molecules. The hydrophobic atom constant for hydrogen was 0.23 for all environments.

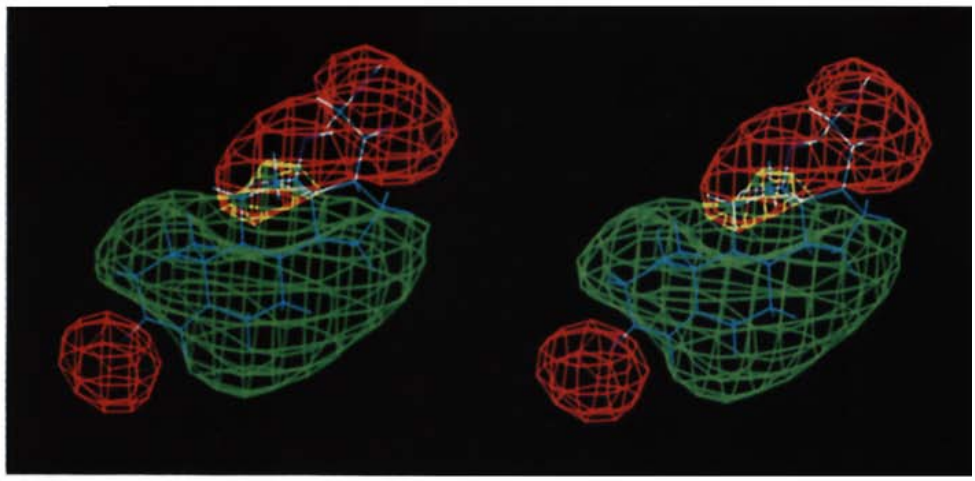
The CoMFA study on steroid binding to corticosteroid-binding globulin was performed according to methods previously published [1] (using SYBYL default values), with the exception of grid spacing. The HINT hydrophobic field was imported into CoMFA with no pre-treatment (e.g., removal of lattice points inside the molecules). In our hands, a 1 Å separation between grid points was found to give more chemically understandable results than larger values, especially for CoMFA studies incorporating the HINT hydrophobic field. This is apparently due, at least in part, to the sensitivity of the hydrophobic field to distance. While the statistical results from grid spacing-dependent studies generally indicate marginally better models with smaller spacing [20], the most important factor is that the resulting CoMFA coefficient contour maps are more interpretable.

EXAMPLE

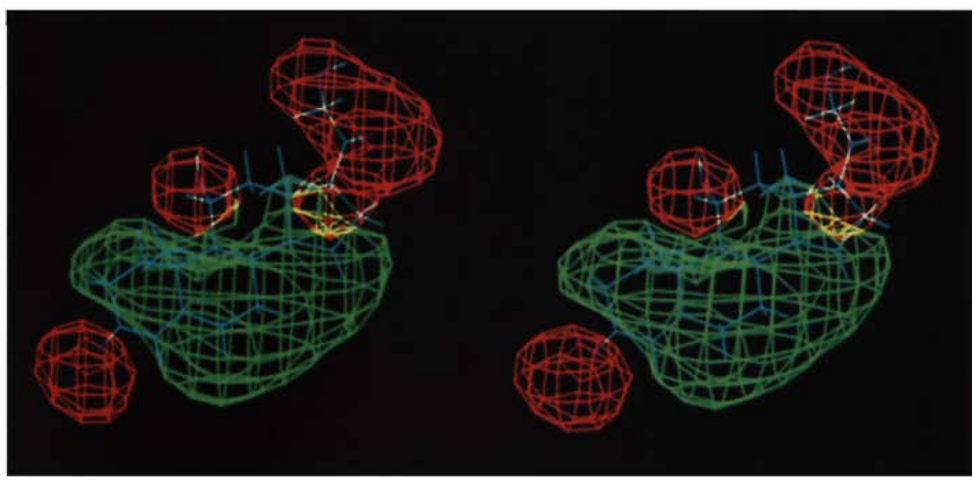
Hydrophobic maps for the 21 steroids were calculated by HINT before analysis with CoMFA. The maps of hydrophobic fields for individual compounds are themselves informative; four are displayed in Figs. 1A–D for the following compounds: aldosterone (A), cortisol (B), cortisone (C), and estradiol (D).



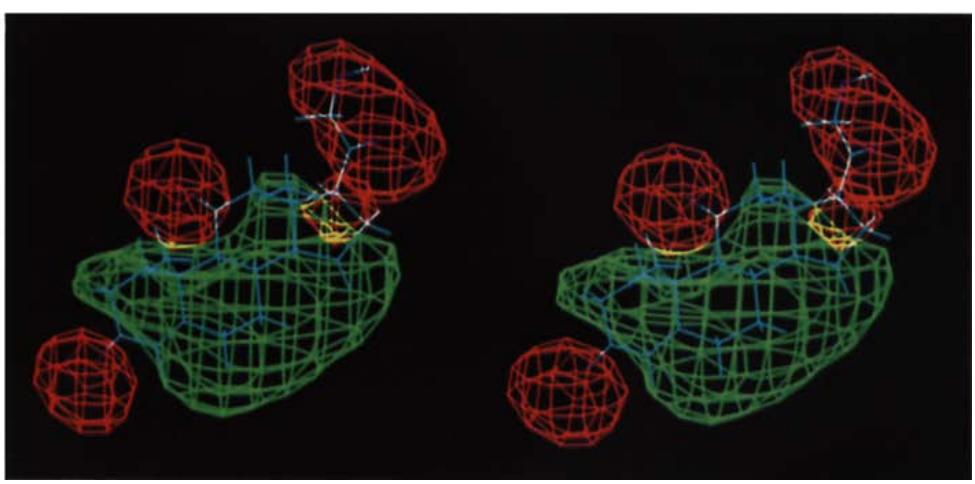
A



B



C



D

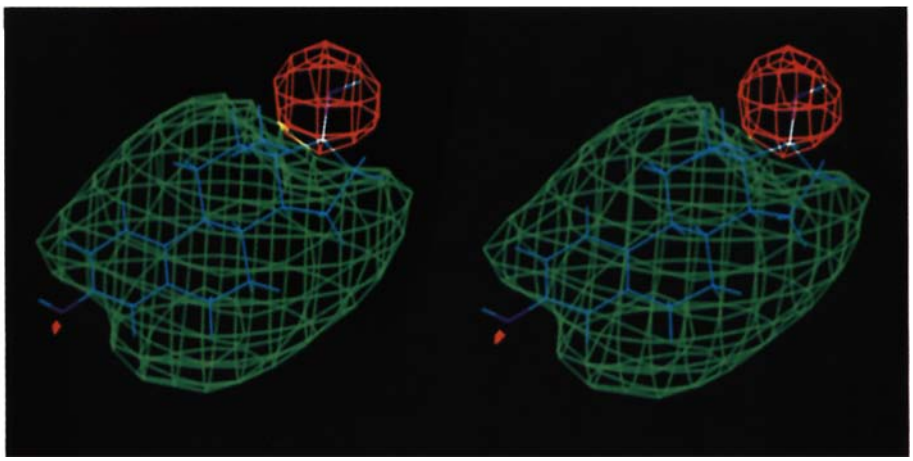


Fig. 1 Hydrophobic field map of (A) aldosterone, (B) cortisol, (C) cortisone, and (D) estradiol. Green contours represent areas of hydrophobicity and red areas represent areas of hydrophilicity. The map is calculated by HINT [12,13] at 1 Å spacing and contoured at −8 (red) and 4 (green) levels. (The apparent yellow contours are an artifact of photographing the overlap of green and red contours.)

The maps, calculated using a 1 Å grid spacing, were contoured and displayed with SYBYL. Hydrophobic regions are shown with green contours while polar (hydrophilic) regions are shown with red contours. The aliphatic (or aromatic A ring in the case of estradiol, Fig. 1D) steroid backbone is clearly depicted as hydrophobic, while the polar hydroxyl, carbonyl, and ether oxygens are easily identified. It is interesting to note that the relative magnitude of hydrophobicity for an atom or fragment is graphically represented by the relative size of the contours encompassing

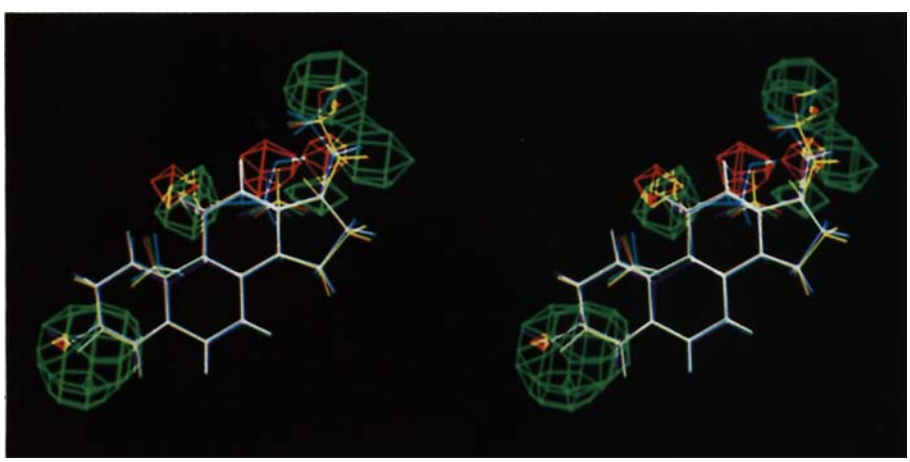


Fig 2. Contoured CoMFA coefficient maps for the hydrophobic contribution to QSAR. The map is calculated from the three-field CoMFA experiment. Green regions, in general, represent molecular modifications where increased hydrophilicity leads to increased biological activity; red regions represent molecular modifications where increased hydrophilicity leads to decreased activity. See text for more information. (The apparent yellow contours are an artifact of photographing the overlap of green and red contours.)

the region. For example, the rather small polar contour for the hydroxyl at position 3 of estradiol is consistent with a decreased hydrophilic character of aromatic substituents as opposed to aliphatic substituents. The contour maps of Figs. 1A–D will be referenced below as aids for interpretation of the CoMFA results.

The steroid data set was examined in CoMFA with three fields (steric, electrostatic, hydrophobic) applied singly and in combinations. Table 1 reports statistical results from these runs. Cross-validated r^2 describes the relative predictive capability of a model that has many more independent variables (grid points) than dependent observables (biological activities). 'Press' is a measure of the difference between predicted and actual biological activity from a cross-validated analysis [15,16]. For the multi-field analyses, the relative contributions of the fields to the cross-validated CoMFA are listed. The cross-validated r^2 values reported in the table are all within the range that suggests reasonable predictive capability.

While examination of Table 1 suggests that, in a statistical sense, inclusion of the HINT field in the CoMFA of steroids does not add to the quality of the model, there are interpretational advantages to including it. The hydrophobic coefficient contour map extracted from the three-field study is shown in Fig. 2. It is important to appreciate that negative hydrophobicity is synonymous with positive hydrophilicity. Thus decreasing hydrophobicity is increasing hydrophilicity, and the color-coded contours in the hydrophobic coefficient maps can be interpreted in terms of either definition. In this example (Fig. 2) all green contour areas indicate where an increase in *hydrophilicity* leads to an increase in activity: hydroxyl substitution at positions 11, 17, and 19; carbonyl substitution at 3 and 18. Red contours are associated with four regions of steroid structural variation. Three of these indicate regions where an increase in *hydrophilicity* leads to a decrease in activity: a phenolic hydroxyl at position 3 (estrogens, Fig. 1D); a carbonyl oxygen at position 11 (cortisones, Fig. 1C); and the cyclic ether and its hydroxyl substituent of aldosterone (Fig. 1A). The remaining red contour can be interpreted as a structural modification where an increase in *hydrophobicity* leads to an increase in activity: the carbon backbone of the C(18), C(19) side chain of the corticosteroids. Significantly, the most active compound in this assay system, cortisol, is defined unambiguously by all of the contour regions of Fig. 2 associated with increasing biological activity.

TABLE 1
COMPARISON OF CoMFA RESULTS FOR STEROID DATA SET FROM FIELD STUDIES

Study	Cross-validated		Conventional		Components	Contribution		
	r^2	'Press'	r^2	s		Steric	Electrostatic	Hydrophobic
Steric only	0.765	0.673	0.974	0.219	2	1.000	–	–
Electrostatic only	0.764	0.674	0.979	0.198	5	–	1.000	–
Hydrophobic only	0.621	0.855	0.667	0.694	1	–	–	1.000
Steric/electrostatic	0.761	0.678	0.904	0.384	2	0.867	0.133	–
	(0.769)[a]				(3)[a]			
Steric/hydrophobic	0.724	0 729	0.972	0.226	5	0.741	–	0.259
Electro./hydro.	0.776	0.657	0.979	0.199	5	–	0.852	0.148
Steric/electro./hydro	0.739	0.709	0.973	0.222	5	0.671	0.108	0.221

[a]Ref. 1.

DISCUSSION

It is interesting that addition of the hydrophobic field to the CoMFA of steroids does not statistically improve the model. However, interpretation of the hydrophobic coefficient contour map as described in the previous section may have clear advantages for drug design over either the steric or electrostatic CoMFA coefficient maps [1]. In this sense, the hydrophobic field 'improves' the model. Inclusion of the HINT field into data sets where activity more strongly correlates with hydrophobicity than does the steroid set may also demonstrate statistical improvement in CoMFA models. We are investigating such cases.

The interpretability of the hydrophobic coefficient contour map (Fig. 2) is enhanced by the fact that the contours are more closely associated with the structural features of the molecules, while the coefficient maps of the standard CoMFA fields are further removed from the molecules. This is apparently due to the relative magnitudes of the raw fields, and the default field cutoff parameter of CoMFA that sets all field values exceeding this parameter to a constant. This difference in field magnitude may also influence the relative field weighting used by CoMFA, which may, in turn, affect the cross-validated r^2 and field contribution to the final CoMFA. User-definable and auto-scaled weighting of the fields and other QSAR parameters as are included in SYBYL version 5.41 will be a worthwhile addition to CoMFA. This would allow examination of the effects of relative parameter weighting on cross-validated r^2, thus permitting derivation of the optimum model.

Because most of the chemical modifications in the steroid data set are oxygen-based (polar and electrostatically negative), the steric, electrostatic and hydrophobic fields show somewhat related information. In general, this would not be the case: the CoMFA electrostatic coefficient map would not be interpretable alone since each field point encodes four bits of information (positive/negative charge and increasing/decreasing activity) with two algebraic signs. Similarly the HINT map encodes hydrophobicity/hydrophilicity and increasing/decreasing activity. In contrast, the algebraic sign in the steric coefficient map *is* associated with activity and can be used to clearly differentiate between the latter two bits of information in each map. When the set of maps is analyzed in tandem, a complete understanding of the information contained in all maps can be obtained. The usefulness of CoMFA as a drug design technique hinges on the ability to interpret CoMFA coefficient maps in terms of their *chemical* information; the addition of the HINT hydrophobic field to the existing steric and electrostatic fields would appear to be a valuable enhancement in this regard.

ACKNOWLEDGEMENTS

We appreciate the assistance of R.D. Cramer, III in importing the HINT field into CoMFA and for helpful suggestions on this manuscript. A portion of this study (the development of HINT) was supported by NIH-HLBI R01-HL-32793. The SYBYL software has been made available through a University Software grant from Tripos Associates.

REFERENCES

1 Cramer, R.D., III, Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.

2 Thomas, B.F., Compton, D.R., Martin, B.R. and Semus, S.F., Mol. Pharmacol., in press.
3 Marshall, G.R., Mayer, D., Naylor, C.B., Hodgkin, E.E., Cramer, R.D., III, In Fauchere, J.-L. (Ed.) QSAR: Quantitative Structure-Activity Relationships in Drug Design, Alan R. Liss, New York, 1989, p. 287.
4 Allen, M.S., Tan, Y.-C., Trudell, M.L., Narayan, K., Schindler, L.R., Martin, M.J., Schultz, C., Hagen, T.J., Koehler, K.F., Codding, P.W., Skolnick, P. and Cook, J.M., J. Med. Chem., 33 (1990) 2343.
5 Norinder, U., J. Comput.-Aided Mol. Design, 4 (1990) 381.
6 Meyer, H., Arch. Exp. Pathol. Pharmakol., 42 (1899) 109.
7 Overton, E., Vierteljahrsschr. Naturforsch. Ges. Zürich, 44 (1899) 88.
8 Hansch, C. and Fujita, T., J. Am. Chem. Soc., 88 (1964) 1616.
9 Rekker, R.F., The Hydrophobic Fragmental Constant: Its Derivation and Application. A Means of Characterizing Membrane Systems, Elsevier Science Publ., Amsterdam, 1977.
10 Hansch, C. and Leo, A.J., Substituent Constants for Correlation Analysis in Chemistry and Biology, J. Wiley and Sons, New York, 1979.
11 Abraham, D.J. and Leo, A.J., Proteins: Struct. Funct. Genetics, 2 (1987) 130.
12 Kellogg, G.E. and Abraham, D.J., Med. Chem. Res., submitted.
13 Wireko, F.C., Kellogg, G.E. and Abraham, D.J., J. Med. Chem., 34 (1991) 758.
14 Fauchere, J.-L., Quarendon, P. and Kaetterer, L., J. Mol. Graphics, 6 (1988) 203.
15 Wold, S., Ruhe, A., Wold, H. and Dunn, W. J., III, SIAM J. Sci. Stat. Comput., 5 (1984) 735.
16 Wold, S., Technometrics, 20 (1978) 397.
17 Kollman, P.A., In Wolff, M.E. (Ed.) Burger's Medicinal Chemistry, Part 1 (The Basis of Medicinal Chemistry), 4th edn., J. Wiley and Sons, New York, 1980, p. 313.
18 Franke, R., Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.
19 Available from Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144.
20 Cross-validated r^2 (press, components) for 2 Å studies otherwise identical to those in Table 1: steric, 0.711 (0.747, 4); electrostatic, 0.694 (0.768, 2); hydrophobic, 0.643 (0.829, 1); steric/electrostatic, 0.744 (0.703, 5); steric/hydrophobic, 0.685 (0.779, 4); electrostatic/hydrophobic, 0.800 (0.621, 3); steric/electro./hydro., 0.794 (0.630, 5).