Amino Acids (1995) 9: 217–228

Amino
Acids


# Modeling the solubility and activity of amino acids with the LCCI method

**L. Pogliani**

Dipartimento di Chimica, Università della Calabria, Rende, Italy

Accepted January 5, 1995

*Because the properties are a consequence of the structure* (Kier and Hall, 1986)

**Summary.** The linear combinations of connectivity indices method (LCCI) is here employed to model the water solubility and activity of 19 natural amino acids. Starting with the molecular connectivity indices, reciprocal and supra molecular connectivity indices are designed to model the solubility and activity spaces of the natural amino acids. The reciprocal and supra molecular reciprocal connectivity indices have been obtained following the variability of the connectivity indices along solubility space of the natural amino acids. A linear combination of the reciprocals of the connectivity indices (LCRCI) showed a satisfactory modelling of the solubility and activity space while a model based on the LCRCI together with the introduction of supra reciprocal molecular connectivity indices for Pro, Ser and Arg achieved an optimal modelling of the solubility and activity space of the natural amino acids.

**Keywords:** Amino acids – Molecular modeling – Linear combination connectivity Indices – Solubility – Activity

## Introduction

Recently, a molecular connectivity model of three physicochemical properties, the crystal densities and the specific rotations of n = 10 and n = 15 natural amino acids respectively and the longitudinal relaxation rates of the α-carbons of n = 8 natural amino acids plus n = 6 cyclic peptides has been presented in this journal (Pogliani, 1994a). The presented model was based on restricted (no more than three indices) homogeneous (or normal or valence connectivity indices) linear combinations of molecular connectivity indices and on the corresponding linear combinations of orthogonal connectivity indices. It was seen that the valence connectivity $\chi^v$ indices were good

descriptors of the crystal densities and relaxation times while the non-valence $\chi$ indices but especially the orthogonal $\Omega$ molecular connectivity indices derived from the corresponding molecular $\chi$ indices were sufficient descriptors of the specific rotations of amino acids. The method of linear combinations of molecular connectivity indices for molecular modeling (LCCI-MM or even LCCI-MC method) in which the used indices are defined in the frame of the widely known Molecular Connectivity theory developed by Randić, Kier and Hall (RKH theory: Randić, 1975; Kier and Hall, 1986; Randić, 1991a,b) has been recently further developed and refined by applying it successfully to different classes of biochemical, organic (saturated and unsaturated) and inorganic compounds (Pogliani, 1993a,b, 1994a,b,c, 1995a,b,c). The LCCI-MM method is able to estimate the physicochemical properties or even the biological activities of molecules in a direct, easy and internal (that is, without introduction of external non-connectivity parameters) way by the aid of structure-explicit topological indices, the connectivity $\chi$ indices, or by the aid of indices directly derived from these last ones as, for example, the orthogonal indices or other special constructions of connectivity $\chi$ indices.

Present study investigates the possibility to model by the aid of linear combinations of connectivity indices or with indices derived directly from them the water solubility and activity of 19 natural amino acids. An original study on the water solubility of n = 13 natural amino acids conducted by the aid of linear combinations of fragment molecular connectivity $\chi_f$ indices, a rather awkward construction of indices based, mainly, on the $\chi$ values of the functional groups of the amino acids, achieved a rather good description of the given solubility space (Pogliani, 1993b). All along this paper we will see how the normal molecular connectivity $\chi$ indices can be used to derive special constructions of connectivity indices the linear combinations of which give a very satisfactory description of the solubility and activity spaces.

## Method

The description of the water solubility of the natural amino acids with the LCCI-MM method starts with the well-known $\delta$ and $\delta^v$ matrix representation of the amino acids (Pogliani, 1993a,b, 1994a and 1995b). In this representation the non-hydrogen atoms of these compounds are substituted by their corresponding delta and valence delta values that are condensed into $2 \cdot n$ matrices. These delta and valence delta values are atom-level numbers that describe the numbers of nearest-neighbours and the number of valence electrons of a non-hydrogen (heteroatom) atom in a molecule respectively. Once the delta matrices of the amino acids are known the molecular connectivity indices used in this study are computed by the aid of the following relations (Kier and Hall, 1986; Pogliani, 1992)

$$D = \Sigma_i \delta_i \tag{1}$$

$$m\chi = \Sigma_p (\delta_1 \ldots \delta_{m+1})^{-1/2} \tag{2}$$

where m = 0, 1, 2, . . . , is the order of the molecular connectivity index and here indices with m= 0 and 1 will be used. Summation in eq. 1 runs over the different delta values of the delta matrices while in eq. 2 it runs over the m-order paths; subscripts 1, 2, . . . , m + 1 are adjacent delta values in these delta matrices. The corresponding valence molecular connectivity indices are obtained with the introduction in eqs. 1 and 2 of the valence delta values of the corresponding $\delta^v$ matrices of the amino acid.

As in our previously cited modeling studies of the physicochemical properties of the amino acids we start our study on the solubility and activity of natural amino acids with a minimal set of connectivity indices as, normally, such minimal sets work very fine (Pogliani, 1995a) in most circumstances

$$\{\chi\} = \{D, D^v, 0\chi, 0\chi^v, 1\chi, 1\chi^v\}$$

Summation in ${}^0\chi = \Sigma_i(\delta_i)^{-1/2}$ runs over the number of heteroatoms and in $1\chi = \Sigma_i(\delta_1\delta_2)^{-1/2}$ it runs over the number of $\sigma$ bonds.

The modelling of the solubility space S and activity space A will be accomplished with the aid of the following dot product, where P = S or A (P stands for property)

$$P = \mathbf{C} \cdot \chi \tag{3}$$

and where **C** is the row correlation vector resulting from the multivariate analysis and $\chi$ is the best connectivity vector, made up of the parameters of the given minimal $\chi$ set plus the unitary connectivity index $\chi^0 \equiv 1$ resulting from the LCCI search. To short-circuit the possibility to obtain negative calculated values, that have no physical meaning, it is better to use a slightly different form of eq. 3, that is, the absolute (Abs) values of the dot product

$$P = \text{Abs}[\mathbf{C} \cdot \chi] \tag{4}$$

Normally, use of this last equation results in a somewhat better modeling of the studied physicochemical property, as we shall see.

The LCCI molecular modeling (MM) method (Pogliani, 1994b, 1995a,b,c) searches the entire space of possible combinations of the given set of $\chi$ indices and sorts the best Q (Q = r/s) combinations together with their F values ($F = f \cdot r^2/v \cdot (1 - r^2)$, f = freedom degrees and v = number of $\chi$ variables). The quality Q factor is used for picking up the best combination of indices as, for a given correlation coefficient r of the regression, it minimises the standard deviation of estimates s, while the F factor controls (for a given r) that the number v of parameters does not grow excessively.

## Results and discussion

In Table 1 the normal connectivity indices for the 19 natural amino acids together with their solubility and activity of the saturated solutions have been collected. Solubilities (g/kg water) and activities (M) calculated from activity coefficients and solubilities at 25°C have been taken from literature (Handbook of Chemistry and Physics, 1991–1992; Christensen, 1994). Clearly, the two spaces, solubility S and activity A are each other related being for a good extent collinear. In fact regressing A against S, the regression coefficient is r(A,S) = 0.92. Even if this is not a strong interrelation in the sense of molecular modeling (Mihalić, Nikolić and Trinastić, 1992, propose a $r > 0.98$ for strong interrelation), nevertheless it indicates the possibility to model both spaces with linear combinations of similar constructions of connectivity indices.

The following are the Q- and F-best combinations of $\chi$ indices that achieve an insufficient modeling of the solubility S and activity A spaces respectively

***S:*** $\{{}^0\chi, {}^1\chi\}$: Q = 0.0022, F = 5.93, r = 0.652, s = 297, n = 19

$\{D, D^v, {}^1\chi, {}^1\chi^v\}$: Q = 0.0023, F = 3.25, r = 0.694, s = 302, n = 19

***A:*** $\{{}^0\chi, {}^1\chi\}$: Q = 0.953, F = 12.6, r = 0.782, s = 0.821, n = 12

$\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi^v\}$: Q = 1.077, F = 6.45, r = 0.844, s = 0.784, n = 19

**Table 1.** Experimental water solubility S (grams per kilogram of water) and Activity A (M) at 25°C of 19 amino acids (AA) and their molecular connectivity index values

| AA | S/A | D | $D^v$ | $0\chi$ | $0\chi^v$ | $1\chi$ | $1\chi^v$ |
|---|---|---|---|---|---|---|---|
| Pro | 1622/5 | 16 | 28 | 5.983 | 4.554 | 3.805 | 2.767 |
| Ser | 422/2.4 | 12 | 28 | 5.862 | 3.664 | 3.181 | 1.774 |
| Gly | 251/2.4 | 8 | 20 | 4.284 | 2.640 | 2.270 | 1.190 |
| Arg | 181/1 | 22 | 42 | 9.560 | 6.709 | 5.537 | 3.600 |
| Ala | 167/1.5 | 10 | 22 | 5.155 | 3.510 | 2.643 | 1.627 |
| Thr | 97/1.5 | 14 | 30 | 6.732 | 4.535 | 3.553 | 2.219 |
| Val | 58/0.45 | 16 | 28 | 6.732 | 5.088 | 3.553 | 2.538 |
| Met | 56/0.4 | 16 | 26.7 | 7.276 | 6.146 | 4.181 | 4.044 |
| His | 43/0.3 | 22 | 42 | 8.268 | 5.819 | 5.198 | 3.155 |
| Gln | 42/0.3 | 18 | 38 | 8.146 | 5.410 | 4.537 | 2.804 |
| Ile | 34/0.27 | 16 | 28 | 7.439 | 5.795 | 4.091 | 3.076 |
| Phe | 29/0.5 | 24 | 42 | 8.975 | 6.604 | 5.698 | 3.722 |
| Asn | 25/0.2 | 16 | 36 | 7.439 | 4.703 | 4.037 | 2.304 |
| Leu | 23/0.17 | 16 | 28 | 7.439 | 5.795 | 4.036 | 3.021 |
| Trp | 12/0.05 | 32 | 54 | 10.836 | 8.104 | 7.182 | 4.716 |
| Glu | 8.6/0.06 | 18 | 40 | 8.146 | 5.280 | 4.537 | 2.739 |
| Lys | 6/0.7 | 18 | 32 | 7.983 | 5.916 | 4.681 | 3.366 |
| Asp | 5/0.04 | 16 | 38 | 7.439 | 4.572 | 4.037 | 2.239 |
| Tyr | 0.5/0.003 | 26 | 48 | 9.845 | 6.974 | 6.092 | 3.857 |

The description of the A space even if it is not satisfactory is more impressive than the description of the S space. The two F-best combinations are the same for both spaces, underlining the partial interrelation of the two spaces. The following $\chi$ and **C** vectors together with eq. 4 have been used to model the solubility and activity spaces respectively

***S***: $\chi = ({}^0\chi, {}^1\chi, \chi^0)$, $\mathbf{C} = (-606.94, 717.2, 1620.1)$

***A***: $\chi = (D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi^v, \chi^0)$,
$\mathbf{C} = (1.0456, -0.6388, 3.7134, -7.9347, 3.6549, 8.2272)$

Eq. 4 has been used as some of the calculated values derived by the aid of eq. 3 are negative, the improvement achieved by eq. 4 over eq. 3 is

***S***: $[\mathbf{C} \cdot \chi]$: Q = 0.0023, F = 12.6; Abs$[\mathbf{C} \cdot \chi]$: Q = 0.0028, F = 19.2

***A***: " Q = 1.23, F = 42.2; " Q = 1.37, F = 52.4

Calculated S values, that are plotted versus the experimental ones in Fig. 1 (S) and 2 (A) confirm the deficient description of both spaces with LCCI. The F-best combination has here been used to simulate the S space as it has not only a better F value but also a better s value while its r value is rather similar to the corresponding value of the Q-best combination; for the A space the Q-best combination has been used as its Q and r values are more impressive.

To gain a deeper insight of the modelling power of the given $\chi$ indices we analyse their variability along the solubility space (the A space being partially collinear behaves in a similar way). In Fig. 3 the experimental water solubility

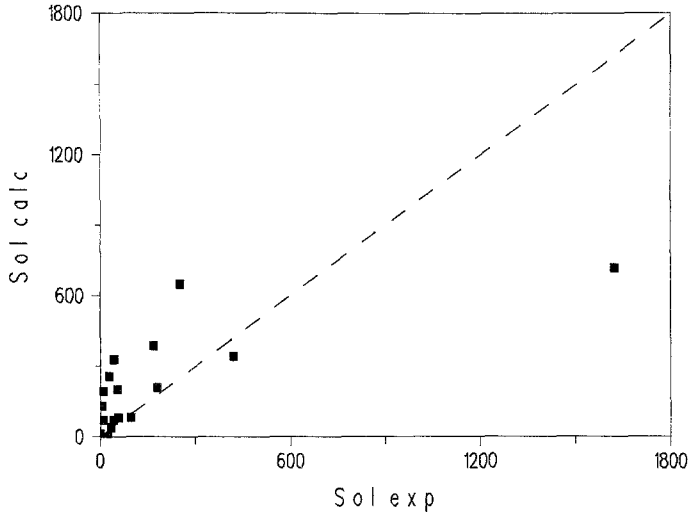


**Fig. 1.** Calculated (by the aid of LCCI and eq. 4) versus experimental solubility of 19 amino acids

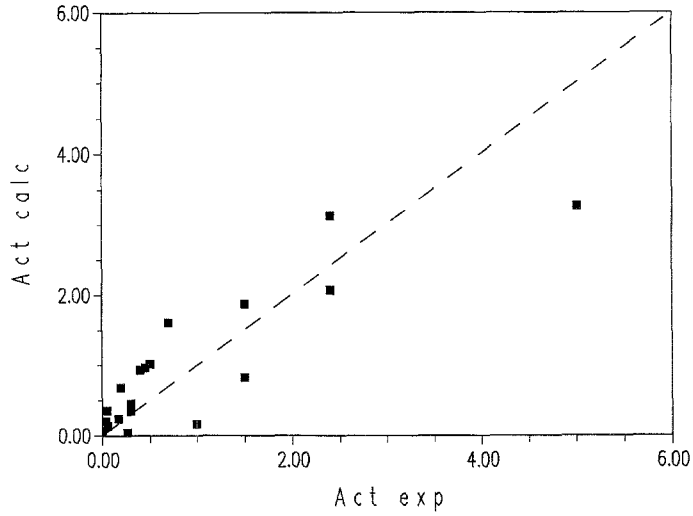


**Fig. 2.** Calculated (by the aid of LCCI and Eq. 4) versus experimental activity of 19 amino acids

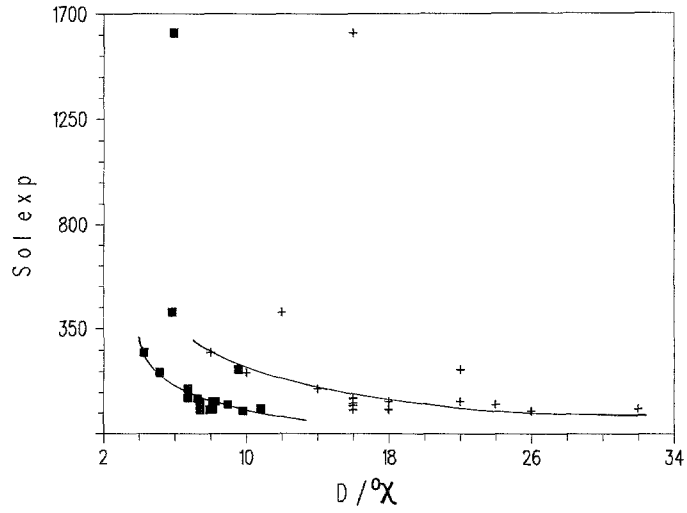


**Fig. 3.** Experimental solubility values of 19 amino acids versus D(+) and $^0\chi$ (full squares) connectivity index values (the solid lines show the hyperbolic variability of the given indices)

values of the 19 amino acids have been plotted versus the pair of connectivity index values D and $0\chi$. The striking feature of this figure is the $\chi$ variability (without outliners Pro, Ser and Arg it is more evident) that shows a hyperbolic-like character, that is, of the type $S \cdot \chi \approx$ cost. This variability, that can be detected also with the other connectivity indices, suggests, then, to introduce the following set of reciprocal connectivity $1/\chi = R$ indices

$$\{R\} = \{{}^{D}R, {}^{D}R^{v}, {}^{0}R, {}^{0}R^{v}, {}^{1}R, {}^{1}R^{v}\}$$

Now, linear combinations of these reciprocal connectivity indices, LCRCI, produce the following description of the solubility and activity space (n = 19)

***S:***
$\{{}^{0}R\}$: Q = 0.0011, F = 2.74, r = 0.373, s = 353
$\{{}^{D}R, {}^{0}R\}$: Q = 0.0037, F = 16.8, r = 0.823, s = 222
$\{{}^{D}R^{v}, {}^{0}R, {}^{1}R^{v}\}$: Q = 0.0044, F = 15.5, r = 0.870, s = 200
$\{{}^{D}R, {}^{0}R, {}^{0}R^{v}, {}^{1}R^{v}\}$: Q = 0.0046, F = 12.8, r = 0.886, s = 194

***A:***
$\{{}^{0}R\}$: Q = 0.622, F = 10.8, r = 0.622, s = 1.0
$\{{}^{0}R, {}^{1}R\}$: Q = 1.22, F = 20.6, r = 0.849, s = 0.697
$\{{}^{D}R^{v}, {}^{0}R, {}^{1}R^{v}\}$: Q = 1.27, F = 15.0, r = 0.866, s = 0.680
$\{{}^{D}R, {}^{0}R, {}^{0}R^{v}, {}^{1}R^{v}\}$: Q = 1.31, F = 11.9, r = 0.879, s = 0.673

The chosen R and C simulating vectors for the two spaces S and A to obtain the calculated values of figures 4 and 5, that are plotted versus the corresponding experimental ones, are

***S:*** $\mathbf{R} = ({}^{D}R, {}^{0}R, {}^{0}R^{v}, {}^{1}R^{v}, R^{0})$,
$\mathbf{C} = (36121, 18636, 22730, -7432.8, -1830.7)$

***A:*** $\mathbf{R} = ({}^{D}R, {}^{0}R, {}^{0}R^{v}, {}^{1}R^{v}, R^{0})$,
$\mathbf{C} = (-89.945, 51.039, 68.454, -21.431, -5.9339)$

From the given series of R combinations we notice that i) the two spaces share the same kind of best combinations, ii) in the S space from the one- to the two-index combination there is a nice statistical improvement, iii) the Q-best combination for both spaces is the fourth one and iv) with the exception of the single-index combinations, the F values, in S and A spaces, of alike combinations are rather similar. The introduction and use of LCRCI seems, then, to be, the right answer to solve the modeling of the solubility and activity spaces of the amino acids within the frame of molecular connectivity Randić, Kier and Hall (RKH) theory. The high values of the standard deviation of the estimate s of these LCRCI are the only negative element, that it will be solved in the next paragraphs.

From Figs. 4 and 5 it is evident the rather satisfactory score of the reciprocal connectivity indices relatively to the normal connectivity indices, from which they have been directly derived. These two figures have also been obtained by the aid the modeling eq. 4 as the statistical scores of both equations are

***S:*** $[\mathbf{C} \cdot \mathbf{R}]$: Q = 0.0050, F = 62; Abs$[\mathbf{C} \cdot \mathbf{R}]$: Q = 0.0072, F = 128

***A:*** " Q = 1.44, F = 57.8; " Q = 1.83, F = 92.6

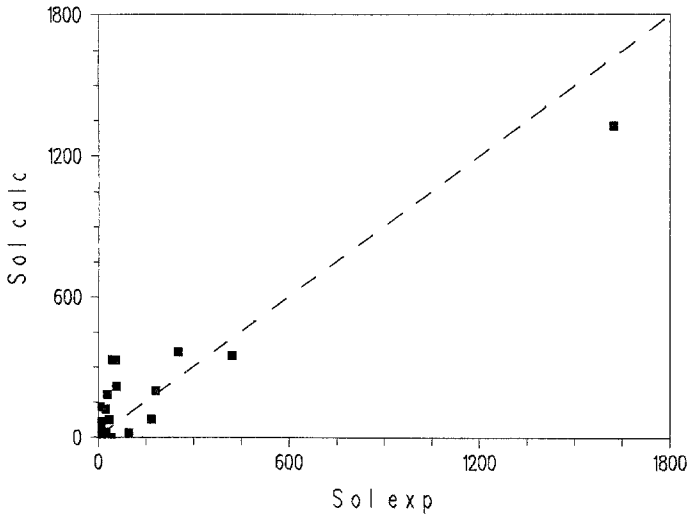


**Fig. 4.** Calculated (by the aid of LCRCI and eq. 4) versus experimental solubility of 9 amino acids

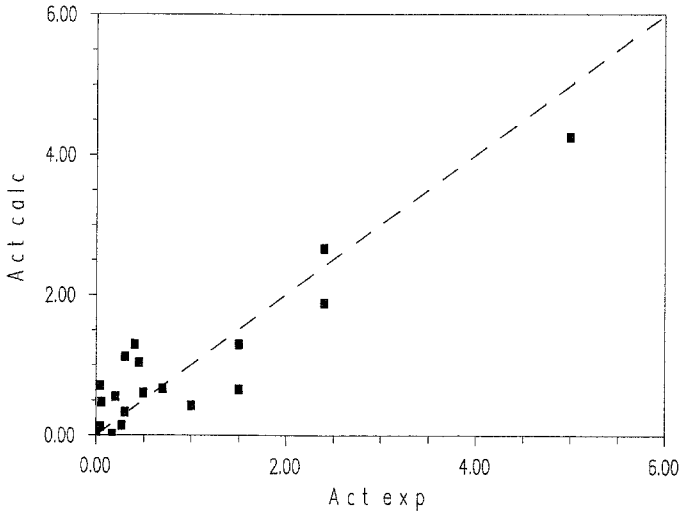


**Fig. 5.** Calculated (by the aid of LCRCI and eq. 4) versus experimental activity of 19 amino acids

Analysing the variability of the connectivity indices all along the solubility space (Fig. 3) we suggested to forget the three high concentration points due to the solubility of Pro, Ser and Arg. Let us model, now, by the aid of LCRCI the n = 16 solubility and activity points. The following are the best successive LCRCI for the description of the two n = 16 S and A spaces

| | | |
|---|---|---|
| ***S:*** | $\{^0R\}$: | Q = 0.038, F = 97.4, r = 0.935, s = 24.7 |
| | $\{^0R, ^1R\}$: | Q = 0.044, F = 66.2, r = 0.954, s = 21.7 |
| | $\{^DR^v, ^0R, ^0R^v\}$: | Q = 0.046, F = 46.8, r = 0.960, s = 21.1 |
| | $\{^DR, ^DR^v, ^0R, ^0R^v\}$: | Q = 0.053, F = 47.9, r = 0.972, s = 18.3 |
| | $\{^DR, ^DR^v, ^0R, R^v, ^1R\}$: | Q = 0.054, F = 39.6, r = 0.976, s = 18.1 |
| | $\{R\}$: | Q = 0.058, F = 37.6, r = 0.981, s = 17.0 |
| ***A:*** | $\{^0R\}$: | Q = 2.58, F = 42.3, r = 0.874, s = 0.339 |
| | $\{^0R, ^1R\}$: | Q = 2.76, F = 25.9, r = 0.894, s = 0.324 |

$\{^0R, {}^1R, {}^1R^v\}$: Q = 2.66, F = 16.1, r = 0.895, s = 0.336
$\{^0R, {}^0R^v, {}^1R, {}^1R^v\}$: Q = 2.57, F = 11.3, r = 0.896, s = 0.349
$\{^DR, {}^0R, {}^0R^v, {}^1R, {}^1R^v\}$: Q = 2.48, F = 8.38, r = 0.899, s = 0.362
$\{R\}$: Q = 2.40, F = 6.55, r = 0.902, s = 0.375

In the following lines we report, instead, the Q-best $\chi$-LCCI combinations for these two n = 16 S and A spaces to have an idea of the improvement introduced by the LCRCI

***S:*** $\{^0\chi, {}^1\chi, {}^1\chi^v\}$: Q = 0.037, F = 31.7, r = 0.942, s = 25.2

***A:*** $\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi^v\}$: Q = 1.08, F = 6.45, r = 0.844, s = 0.784

Comparing these last Q-best LCCI with the preceding Q-best LCRCI for the two n = 16 S and A spaces we notice at once the better Q and F score of the LCRCI. For these two spaces the single-index LCRCI shows already a good statistical score, and, while for the S space the successive combi-

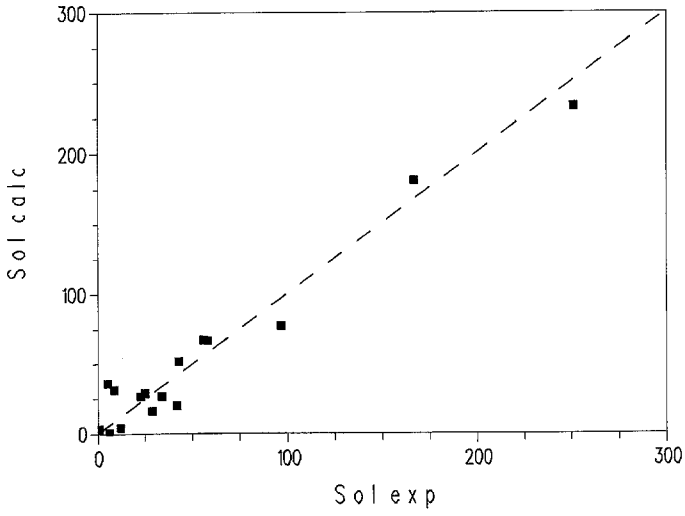


**Fig. 6.** Calculated (by the aid of LCRCI and eq. 4) versus experimental solubility of 16 amino acids

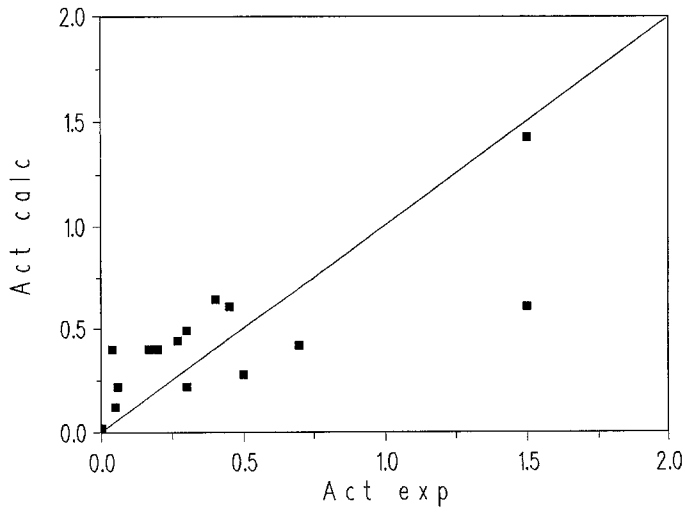


**Fig. 7.** Calculated (by the aid of LCRCI and Eq. 4) versus experimental activity of 16 amino acids

nations get better and better, for the A space the best combination is the second one with Q = 2.76 and F = 25.9. To derive the calculated S and A values that are plotted versus the corresponding experimental ones in Figs. 6 and 7 for this n = 16 space the following **R** and **C** vectors together with eq. 4 have been used

***S:*** **R** = ($^{D}R$, $^{D}R^{v}$, $^{0}R$, $^{0}R^{v}$, $R^{0}$),
**C** = (8113.1, −26638, 11696, −5227.3, −199.08)

***A:*** **R** = ($^{0}R$, $^{1}R$, $R^{0}$), **C** = (44.931, −12.617, −2.5135)

Even here eq. 4 describes better than eq. 3 the two spaces

***S:*** Eq. 3: Q = 0.060, F = 243.7; Eq. 4: Q = 0.062, F = 262.7

***A:*** " Q = 2.86, F = 55.8; " Q = 2.91, F = 57.7

To model the n = 16 S space the 4-R-index combination has been chosen as, even if it does not show the best Q value, it shows nevertheless a better F value than the following two combinations.

To model the very high solubility and activity values of Pro, Ser and Arg, that is, to model the full solubility and activity n = 19 spaces let us now to introduce for these amino acids the supra molecular reciprocal or supra-reciprocal connectivity indices, $a \cdot (1/\chi) = a \cdot R$ (where $a > 1$ can be considered an association parameter), as recently done for caffein homologues, where supra-connectivity $a \cdot \chi$ indices have been successfully used (Pogliani, 1995a). There is something unusual in the high solubility and activity values of Pro, Ser and Arg and to appreciate this anomaly it would suffice to compare their S and A values with the corresponding values of three similar amino acids: His, Thr and Lys. Even if there is no experimental evidence that these three amino acids undergo intermolecular association in aqueous solutions, their very high S and A values cannot be understand without supposing self-association or strong solvation phenomena. These solvation effects can also be simulated by the aid of supra molecular connectivity indices (Pogliani, 1993a and 1995a) formally similar to the proposed indices for self-association. Utilisation of LCRCI that include supraconnectivity R indices with a = 4, 1.5 and 2 for Pro, Ser and Arg respectively, give the following best (single-, Q- and F-best) descriptions of the full S and A spaces

***S:*** {$^{0}R$}: Q = 0.022, F = 1141, r = 0.9926, s = 46.1
{$^{0}R$, $^{1}R$}: Q = 0.029, F = 1028, r = 0.9961, s = 34.4
{$^{D}R^{v}$, $^{0}R$, $^{1}R^{v}$}: Q = 0.030, F = 732, r = 0.9966, s = 33.3

***A:*** {$^{D}R$}: Q = 2.82, F = 222, r = 0.964, s = 0.341

The description of both full spaces is now optimal and while the best S space simulation is achieved by a three-index combination the best simulation of the A space is a single-index simulation. Equation 4 and the following **R** and **C** vectors have been used to model the full S and A spaces

***S:*** **R** = ($^{D}R^{v}$, $^{0}R$, $^{1}R^{v}$, $R^{0}$), **C** = (−8018.9, 5271.0, −316.16, −299.06)
Eq. 4: Q = 0.034, F = 2795; Eq. 3: Q = 0.032, F = 2488

*A*: $\mathbf{R} = ({}^{D}R, R^{0}), \mathbf{C} = (24.295, -0.96039)$
Eq. 4: Q = 2.87, F = 229; Eq. 3: Q = 2.82, F = 222

The good quality of the found modeling for the two spaces can be seen in Figs. 8 and 9.

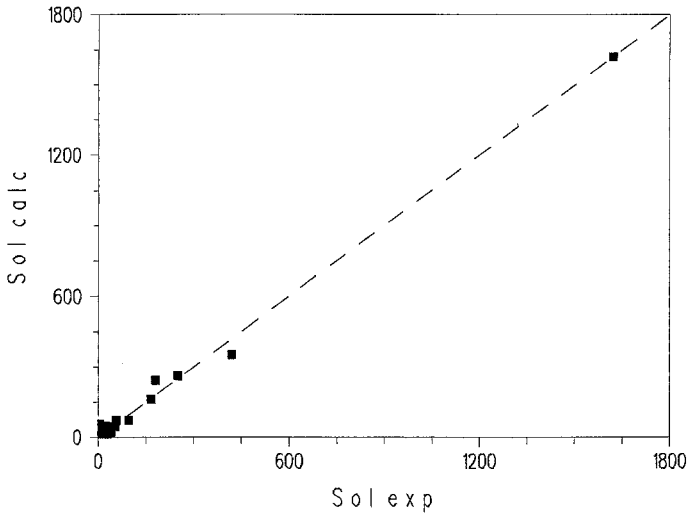


**Fig. 8.** Calculated (by the aid of eq. 4 and of LCRCI plus supraconnectivity reciprocal indices for Pro, Ser and Arg) versus experimental solubility of 19 amino acids

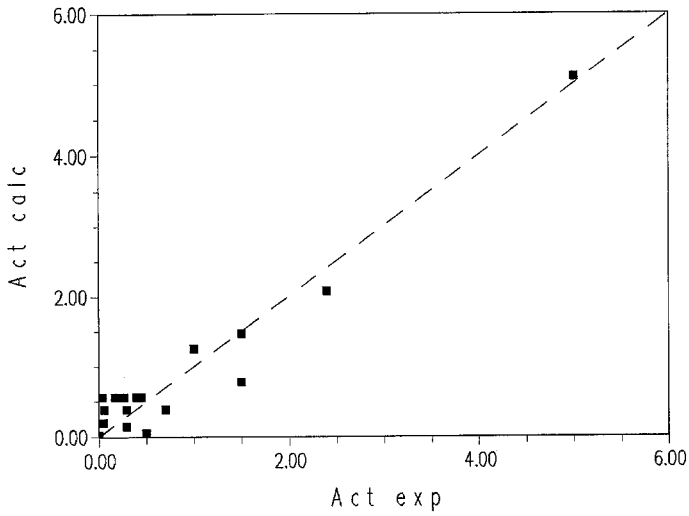


**Fig. 9.** Calculated (by the aid of Eq. 4 and of LCRCI plus supraconnectivity reciprocal indices for Pro, Ser and Arg) versus experimental activity of 19 amino acids

## Conclusion

A minimal set of 6 graph theoretical indices defined within the frame of the RKH molecular connectivity theory is here used to derive, by the aid of an analysis of their variability along the solubility space, new kind of theoretical indices, the reciprocal molecular connectivity indices. These indices achieve a quite satisfactory modeling of the activity and solubility values of n = 19 natural amino acids. The modeling improves if the high solubility and activity

values of Pro, Ser and Arg are left out and the simulation is restricted to n = 16 natural amino acids. To achieve an optimal modeling of the full n = 19 A and S spaces supra reciprocal connectivity indices have to be introduced for Pro, Ser and Arg to take into account their much greater solubility and activity. These supra indices are not experimentally grounded, they are just inferred on the basis of rational arguments that explain these abnormally high S and A values as a result of self-association or strong solvation phenomena. An important aspect of this modeling is the better descriptive power of equation 4 than equation 3. Furthermore, use of this equation allows to handle with positive calculated values excluding negative values that have no physical meaning. To notice is the fact that even if the two spaces, S and A, are to some extent collinear their optimal modeling is normally done by the aid of different combinations of R indices. An optimal description of the two spaces should allow us to derive the activity coefficients (generally an arduous task), that relate concentration to activity, of compounds not included in the modeling but the $\chi$ or R values of which can be easily derived. These activity coefficients, for concentrated solutions are normally obtained by the aid of approximate expressions that normally give no satisfactory or even semiempirical explanation of the data.

## Acknowledgements

This paper is dedicated to the twenthief anniversary of the RKH molecular connectivity theory developed by Randić, Kier and Hall.

## References

CRC Handbook of Chemistry and Physics (1991–1992) 72nd edn. Boca Raton, pp: 7-1

Christensen U (1994) Kinetic characterization of carboxypeptidase-Y-catalyzed peptide semisynthesis prediction of yields. Amino Acids 6: 177–187

Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Wiley, New York (and references therein)

Mihalić Z, Nikolić S, Trinajstić N (1992) Comparative study of molecular descriptors derived from distance matrix. J Chem Inf Comput Sci 32: 28–37

Pogliani L (1992) Molecular connectivity model for determination of isoelectric point of amino acids. J Pharm Sci 81: 334–336

Pogliani L (1993a) Molecular connectivity model for determination of $T_1$ relaxation times of α-carbons of amino acids and cyclic dipeptides. Comput Chem 17: 283–286

Pogliani L (1993b) Molecular connectivity model for determination of physicochemical properties of α-amino acids. J Phys Chem 97: 6731–6736

Pogliani L (1994a) Structure property relationships of some amino acids and dipeptides. Amino Acids 6: 141–153

Pogliani L (1994b) Molecular connectivity descriptors of the physicochemical properties of the α-amino acids. J Phys Chem 98: 1494–1499

Pogliani L (1994c) On a graph theoretical characterization of cis/trans isomers. J Chem Inf Comput Sci 34: 801–804

Pogliani L (1995a) Molecular modeling with linear combination of connectivity indices. J Phys Chem 99: 925–937

Pogliani L (1995b) The molecular connectivity method: a powerful tool in the study of biological relevant molecules. Curr Top Pept Prot Res (in press)
Pogliani L (1995c) Modeling hydration processes of biochemical and inorganic compounds. (submitted)
Randić M (1975) On characterization of molecular branching. J Am Chem Soc 97: 6609–6615
Randić M (1991a) Orthogonal molecular descriptors. N J Chem 15: 517–525
Randić M (1991b) Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. J Chem Inf Comput Sci 31: 311–320

**Author's address:** Dr. L. Pogliani, Dipartimento di Chimica, Università della Calabria, I-87030 Rende (CS), Italy.

Received November 1, 1994